PROPOSAL OF AN EVALUATION METHOD OF A COMPACT CITY MODEL

by

MANRIQUE RUIZ LUIS CARLOS

Department of Social Intelligence and Informatics
Graduate School of Information Systems
The University of Electro-Communications

Dissertation submitted in partial fulfillment of the requirements
for the Degree of
Doctor of Philosophy

September 2014

PROPOSAL OF AN EVALUATION METHOD OF A COMPACT CITY MODEL

APPROVED BY SUPERVISORY COMMITTEE

Chairperson:	Associate professor Kayoko Yamamoto
Member:	Professor Kazuyuki Suzuki
Member:	Professor Satoshi Kurihara
Member:	Associate professor Hisahi Koga
Member:	Professor Michiko Tsubaki

. . . I can appreciate the beauty of a flower. At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes. The fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement, the mystery and the awe of a flower. It only adds. I don't understand how it subtracts.

— Richard Feynman

Dedicated to my family.

其疾如風
其徐如林
侵掠如火
不動如山

武田信玄

論文の和文要旨

都市計画学、建築学、土木工学、環境科学、社会学、経済学などの多様な学問分野において、都市における生活の質の向上を図るために、エネルギーや自然資源等の利用を削減するための新しい考え方が提案され、実現されている。近年では世界中で持続可能な発展のための都市モデルとして、グリーンシティ、スマートシティ、コンパクトシティなどが実現されつつある。これらの都市モデルのうちでも「コンパクトシティ」は、1975 年に建築家 G.B.ダンツィヒと T.L.サティ（アメリカ）により、効率性が高い都市モデルとして提案された。1990 年代以降はコンパクトシティに関する議論が世界各地でなされ、ヨーロッパ諸国ではヨーロッパの都市全体における都市戦略として位置付けられ、世界環境会議（リオネジャネイロ、1992 年）でも持続可能な発展のための重要な都市モデルとして提案されていた。

日本では阪神・淡路大震災（1995 年）の復興のために目指すべき都市モデルとして、神戸市が公式に政策として導入したことを契機として注目されるようになった。日本型コンパクトシティの場合には、持続可能な発展に加えて、超高齢化・人口減少といった深刻な社会問題への対応を考慮して、中心市街地の活性化、都市郊外化・スプロール化の抑制のための都市再生や、市街地のスケールを小さく保ち、歩いて行ける範囲を生活圏と捉え、コミュニティの再生や住みやすいまちづくりを目指すことを目的としている。近年ではコンパクトシティを政策として導入する地方都市が増加しており、冬季に積雪量の多い北海道や東北地方では、市街地の拡大による除雪費用増大の問題の解決も目的に含まれている。

以上の背景に基づき、本研究はコンパクトシティの 2 つの基礎的条件を基に、第 1 段階では人口密度分布と土地利用との関連性、第 2 段階では社会的・経済的要因と土地利用との関連性に着目した評価方法を提案し、第 3 段階ではこれらの評価方法を統合した評価方法を提案した。本論文は、以下のように 6 章から構成されている。

第 1 章では、研究の背景として、これまでに提案されてきた都市モデルについて紹介するとともに、先行研究の成果を引用しつつ「コンパクトシティ」の都市モデルとしての重要性について提示した上で、本研究の目的を明示した。

第 2 章では、コンパクトシティの定義と日本型コンパクトシティの特徴について整理した上で、日本のコンパクトシティに関連した都市計画制度の概要について整理し、これらを基に本論文全体の評価の枠組みを構築した。さらに本論文の評価対象地域として、日本でコンパクトシティを政策に導入している都市のうち、自然的・社会的条件からコンパクトシティ実現の必要性が最も高いが、規模の異なる 3 ヶ所の地方都市圏（札幌都市圏、仙台都市圏、青森都市圏）を選定した。また本論文における基礎データについて紹介した上で、地理情報システム（GIS: Geographic Information Systems）等を用いて、本論文で利用可能なシェイプファイル形式にデータを加工する方法について説明した。さらに、本論文で用いるソフトウエアについても説明を加えた。

第 3 章では、第 1 段階の評価として、コンパクトシティの基礎的条件のうち、人口密度分布と土地利用との関連性に着目した評価方法を提案した。評価方法として、

Residual Kriging Model の Compactness と Shannon の Entropy を指標として用いて、人口密度と市街地のコンパクト性と拡散性について評価を行う方法を提案した。この評価方法を 3 ヶ所の評価対象地域に適用し、利用可能なデータが入手できた 1991 年、1997 年、2006 年の 3 時点で評価を行い、各都市圏における結果について考察した。

第 4 章では、第 2 段階の評価として、コンパクトシティの基礎的条件のうち、社会的・経済的要因と土地利用との関連性に着目し、Support Vector Machine を用いた評価方法を提案した。社会的・経済的要因としては、交通施設、公的施設、医療施設、商業施設等の日常生活で必要な諸施設への近接性を取り上げた。第 3 章と同様にこの評価方法を 3 ヶ所の評価対象地域に適用したが、社会的・経済的要因に関する過去のデータが入手不可能であったため、最新の 2006 年のデータのみを用いて評価を行い、各都市圏における結果について考察した。

第 5 章では、第 3 段階の評価として、第 3 章と第 4 章で提案した評価方法を統合した評価方法をさらに提案した。前 2 章と同様にこの評価方法を 3 ヶ所の評価対象地域に適用し、最新の 2006 年のデータを用いて評価を行い、各都市圏における結果について考察した。また都市計画学、建築学、土木工学、環境科学等の専門分野における 12 名の専門家にヒアリングを行い、本論文で提案した評価方法の妥当性を確認するとともに、今後考慮すべき点について助言をいただいた。

最後に、第 6 章では、まず本論文全体のまとめを行い、結論を総括した。今後の研究課題として、第 5 章における専門家からの助言を考慮し、エネルギーの効率性、交通量、建物の容積率等を考慮した評価方法が必要であることを提示した。

ABSTRACT

The urban planning has always been a challenge for the humankind. Experts in academic fields such as urban planning, architecture, civil engineering, environmental science, sociology, economy and others have been implementing and proposing new ideas to minimize costs, energy and different resources in order to improve the quality of life in the cities. Models looking for sustainability are being applied in different cities around the world, for instance models such as green cities, smart cities and compact cities. In Japan the compact city model is implemented in different cities since it suits the natural and geographical conditions; to evaluate the implementation of this model, there are different techniques that take into account several type of variables such as population, land use and some using socio-economic factors. However, these evaluation models lack the integration of all the aspects before mentioned. For that reason, this study develops an evaluation method of a compact city model combining all the abovementioned variables. Thus, the doctoral dissertation is divided in 6 main characters explained as follows:

Chapter 1 introduces the necessity and purpose of an evaluation of a compact city model and the background of this study. Chapter 2 shows the compact city model definition in the world and its application in Japanese urban planning, it provides the framework and outline of evaluation methods, and study areas description. This chapter also introduces the software used for the study to manipulate the different databases, and elaborate the mathematical models and Geographic Information Systems (GIS).

Chapter 3 addresses the evaluation focusing on population and land use, methodology, indexes calculations and results. In this chapter a residual Kriging model is applied using population density to describe the urban settlement in 3 metropolitan areas in 3 different periods of time. Chapter 4 addresses the evaluation focusing on socioeconomic factors and land use. As Chapter 2, the methodology is provided paying particular attention to support vector machine calculation and its implementation using parallel computing to predict land use in 3 metropolitan areas; later the area under the curve and cross validation methods are implemented to calculate the most significant socio-economic factors for each type of land use.

Chapter 5 gathers the most important results and indexes to develop the evaluation method of a compact city model using information related to population, land use and socio-economic factors. In this chapter points of view of chemistry and thermodynamics are applied so as to evaluate the relation between the entropy and com-

pactness. An atomic structure reflecting the relationship among the variables is also provided. Finally, the evaluation method of the compact city model and its validation in 3 different metropolitan areas of Japan is offered.

Chapter 6 concludes all the results in the study, and shows the future work in urban planning, implementation of different geo-statistical models and hardware constrains to implement different systems.

ACKNOWLEDGMENTS

First, I would like to thank to God for giving me health and a beautiful family.

To Sandra Puentes for all your assistance, help and advices, for being my lovely wife and partner in science.

To Prof. Yamamoto to teach me about city planning and expecting always the best of me.

To my family and friends for all your kind comments and support. And also special thanks to Neil Rubens and Shiga Mikio for your kindness, priceless comments and suggestions.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AOI Area of Interest

AUC Area Under the Curve

GIS Geographic Information System

LUCKY Land-Use Control Back-up System

MLIT Ministry of Land, Infrastructure, Transport and Tourism

MtA Metropolitan Area

MtAs Metropolitan Areas

OSH Optimal Separated Hyperplane

ROC Receiving Operating Characteristics

SVM Support Vector Machine

UCA Urbanization Control Area

UPA Urbanization Promotion Area

1 INTRODUCTION

In contrast with European and American countries, most large cities in Asia have a high population density [98]. It is important for each government and city planners to consider how to improve living conditions; because poor planning results in more problems such as environmental degradation, disease, lack of green spaces for diverse functions like recreation, environmental conservation, disaster prevention and unintended changes to the local topography. According to Yamamoto, Japan lacks open green spaces compared to other large cities in advanced countries [97]. Therefore, it is important for scientists, urban planners and architects to consider how to create or improve new kinds of city planning models such as compact, smart, green cities, etc.

This study focus on the compact city model; because it is one of the most important urban designs applied to cities in recent years, focusing mainly on population density, socio-economic factors, central area revitalization, mixed-use development, services and facilities. For instance, Oslo, Groningen, Amsterdam; and in Japan: Sapporo, Wakkanai, Sendai among other cities are examples of the application of the compact city model around the world. Because of the previous reasons, a proposal of an evaluation method for a compact city model taking in count variables such as population, land use and socio-economic factors is provided. In the first part, I will implement a mathematical method to evaluate the land use according to the master plan of the different Metropolitan Areas (MtAs) in different periods of time. With this information it is possible to calculate some indexes for evaluating the application of the compact city model.

In the second part, I will use socio-economic factors to infer how they affect the land use; developing a mathematical model I will determine which socio-economic factors have a major impact on a specific type of land use. Finally, I will gather the most important results to develop the evaluation method taking into account the factors previously mentioned. Through this method I gather data related to population density, land-use and socio-economic factors to evaluate appropriately a Japanese compact city model. Thus, I will focus on two main aspects of the compact city model: The first one is related to population distribution, and the second one related to socio-economic factors and how they affect the land use.

This document provides academic approaches such as the application of the different methods in this particular study and the use of data appropriately. The different MtAs vary in size and this makes more complex calculations, for that reason it was necessary to use adequately the hardware resources so as to reduce the calculation time especially for the Support Vector Machine (SVM) model application. As a social implications, 3 periods of time were evaluated to describe the population settlement in the different MtAs; these experiments were useful to identify the urban core areas as well as the commuter belts.

In order to identify the position of this study in the literature, allow me to describe the state of art in the related fields. Scientists are developing different methods to predict geographical information in various academic fields. Methods applied to social physics [89], agriculture [32], climatology [92] and those related to decreasing population density [13] are examples of their broad coverage. Some models follow the probability theory, and predictions are accompanied by estimates of prediction errors. Those models can be grouped based on: Smoothing effect, proximity effect, convexity effect and support size. The latter one makes predictions at points or for blocks of land, they are called area-based and point-based [36]. Point-based methods, which depend initially on geographic coordinates, use zone centroids as control points. Later, a window is positioned over each control point and the source zone population is allocated to grid cells, falling inside this window using unique weighting based on the distance decay function between the source zone centroid and the grid cell. On the other hand, area-based methods use the zone itself as the unit of operation; this method is more concerned with volume preservation [48]. However, if the boundaries are not symmetrical, the location of the centroid can be affected, and as a result interpolation may be biased [96]. For this reason, this study works with 100m mesh to provide more accuracy than the original extent.

Some methods have been developed to handle point-based methods, for instance, interpolating polynomials, Kriging, distance - weighting methods, etc. For Kriging methods, if the source zones can be reduced to point sources and the population distribution can be described with a semi-variogram [30], this method becomes the best linear unbiased estimation [51]. The Kriging model is also known as BLUP (Best Linear Unbiased Predictor), and is used in diverse studies [32] [88]. In order to calculate the best predictor for unknown quantities, it is important to calculate the best linear unbiased predictor [8]. The study developed by Longsdon et al.[52] focused on the probability mapping of land use change using a Geographic Information System (GIS) interface for visualizing transition probabilities. Related to population, Araki [4] analyzes the change in population and land-use intensities in Namibia; and the studies of Martin [56], who developed a kernel-based interpolation algorithm to ensure that the populations reported for target zones are constrained to match the overall sum of the source units.

The studies related to socio-economic factors were also addressed. For instance, the studies such as that elaborated by Kawata [44] analyzing the socio-economic factors influencing longevity in Japan, and the study elaborated by Fukuda [24] focusing on the municipal socio-economic status and mortality in Japan. Also, the studies related to compact city models around the globe were considered; for example the study by Denpaiboon [16] describing a compact city strategy of Bangkok mega city; and the study by Burton [11] measuring the urban compactness in towns and cities in the U.K. using land use information. The term compactness has been defined by different authors [31] [20] [29], etc., all of them have in common the high-density or monocentric development, concentration of employment, housing and mixture of land uses. However, Tsai[93] found a common theme that involves the concentration of development. Another important indicator is calculated to measure the urban sprawl; it is named as Shannon's entropy for land use classification. The urban sprawling brings a complex situation in different countries; and there is not an unified meaning [91]. Based on the idea that there are different types of land use in a region of interest, it is possible to apply this concept with GIS. It has been proven that Shannon's entropy index is an efficient approach to the measurement of urban sprawl[82].

The SVM is a supervised non-parametric statistical learning technique, which has an important property: The determination of the model parameters corresponds to a convex optimization problem, where a local solution is also a global optimum. Recently, SVM is applied in different kind of studies. For instance Zhou et al. [103] presented a method of Japanese dependency structure analysis based on SVM and conditional random fields (CRF). Their experiments demonstrated that combining SVM and CRF outperforms the cascaded chunking model based on sole SVMs and sole CRFs. Sudha and Bhavani [90] compared the efficiency between the k-Nearest Neighbor models (kNN) and multi class SVMs where multiple gait components were fused for enhancing the classification rate. Their results demonstrated that the classification method using SVM was better than kNN. Pitiranggon et al. [69] developed a decompositional rule extraction technique from SVM called Support Vector Space Expansion (SVSE) rule. They applied it to financial data to predict currency crises. Ramirez-Gutierrez et al. [75] developed a face recognition algorithm using eigenphases and histogram equalization. They proposed a featured extraction scheme using SVM, the recognition rate was higher than 97% and verification error lower than 0.003%. Agouris et al. [2] described a procedure integrating spatial coordinates to data by developing a SVM, they scaled point attributes in order to be in the same domain type, later they implemented a chunking option for working with large training data. Also the approach by Yu [102], an SVM was developed using a ranking function in order to find a house in Chicago.

Relevant studies related to land use are in remote-sensed data. Plaza et al.[70] focused on the methodologies for processing a specific type of imagery using SVM. Zhu and Blumberg [104] analyzed the different results of satellite image data, applying different kernels such as radial basis and polynomials. Foody and Mathur [22] studied the potential for intelligent training sample collection; it was applied in classification of agricultural crops from multispectral satellite sensor data. They could classify crops with 92.5% of accuracy. Provost and Fawcett [71] discussed the importance of the area under the Receiving Operating Characteristics (ROC) curve called the Area Under the Curve (AUC). They developed a hybrid classifier for any target conditions; the model is based on a method for comparison of classifier performance. This is done by combining techniques from ROC analysis, decision analysis and computational geometry. Brefeld and Scheffer[9] discussed an approximation for large datasets that clusters the constraints. Developing an AUC maximizing kernel machine they optimized a bound on the AUC and a margin term.

Because of the problems related to settlement not only in Japan, also in different parts of the world, it is required to use appropriately the natural resources and land use in order to prevent disasters. In Japan, for example, there are problems related to depopulation and aging where the life expectancy is 83 years old; for that reason it is necessary to gather the residents in central parts of cities. Another important constraint is the related to the habitable area, because it is less than 21% of its landmass, while the 66% corresponds to forest [42]. Consequently, a compact city model is necessary to improve the quality of life of individuals, preservation of the environment and proper use of economic and technological resources.

The originality of this study relies on the integration of the different models since other authors concentrate their efforts just in the analysis of the compact city model in the land use system, population or in the master plan; however there is a lack of integration between these elements and socio-economic factors. This is due to the complexity to gather different kind of information and its processing. In order to be more accurate with the results, I have used the most detailed information provided by the Japanese government, also the acquisition of the socio-economic factors data has been done exhaustively. Concepts of chemistry and thermodynamics have been applied to evaluate the compact city model in the MtAs. Finally, conclusions and future work give an overview of the application of the previously mentioned scientific fields in the future of city planning pursuing the compact city model. Considering that a Compact city model depends on specific characteristics such as geography, laws, master plans among other aspects; this proposed evaluation method can be applied in any particular case contemplating all the factors above mentioned.

2 FRAMEWORK AND DEFINITIONS

This chapter addresses the compact city model, definition and some approaches around the globe. It will explain the special characteristics of the Japanese compact city model, focusing on the necessary objectives to reach this goal; it also explains the legal city planning system in Japan. The framework is divided in four main parts. The first part explains the compact city model and its characteristics. The second part explains the legal city planning system in Japan and how the different laws have been implemented by the local governments.

Description of study area will explain why these is metropolitan areas (MtAs) should be studied; it will explain the master plan of them and their differences in the land use. The collection and processing of population, land use and socio-economic factors data is explained. Population data explains how this information was divided, and provide some statistics in the different periods of time noticing the merging of municipalities and the problem of depopulation. Land use will explain in detail how the government defines the different types of land use. Socio-economic factors explain why and how they were chosen.

Finally the software tools used to develop this project will be introduced.

2.1 COMPACT CITY

2.1.1 *Definition*

The term of compact city was first coined by Dantzig and Saty in 1973 [15]. The main goal is reducing distances between the personal, social and economic activities of citizens. The compact city model focuses on population density, activity concentration, public transportation intensification and city size and access conditions [79] [53].

In the early 1980s, K. Lynch suggested a model where cities resemble a galaxy of separate medium-sized communities, surrounded by large swaths of green space and connected by major roads [55]. He mentioned that the compact city should be located as part of a distinct regional area, with clear boundaries designated to ecological and social functions. Compact city strategies also include a large block or open space close to urban neighborhoods, and the intensification of abandoned or unused land within the urban core. The large blocks or superblocks present two patterns such as high and low buildings, and wide roads and narrow streets. Major roads divide the city into a

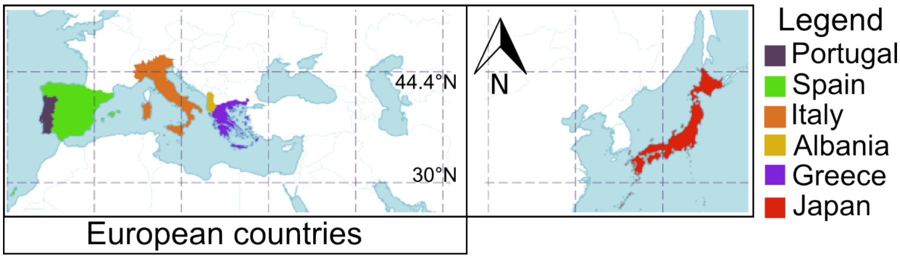


Figure 2.1: Comparison between European countries and Japan

series of 'superblocks' and within these blocks is a network of narrow streets made up essentially of streets without sidewalks. In this way, the road grid can be called a superblock grid [83].

One of the first compact city models was developed by Swiss architect and urban designer, LeCorbusier [14]. He thought that each city should be concentrated with high-density urban living associated with high-rise residential buildings. Later, in 1994, Haughton and Hunter described a concentrated city model, explaining the importance of improving urbanization, industrialization and housing policies among other factors [34].

The concept of compact cities around the world depends on the physical attributes of each area. Thus, it is possible to compare Japanese compact cities with European ones, since both regions have limited space, and housing, businesses, hospitals, etc., are located in specific areas according to a master city plan. In contrast, cities in the U.S. have a large area available for housing and industry.

2.1.2 *Characteristics of a Japanese compact city*

In northern Japan, cities must deal with snow-related problems every year and invest a part of their budget to control this situation. Another constraint is that the habitable area of Japan is less than 21% of its landmass, and 66% is forest [42]. As a result, urban sprawling is not as great an issue as in other countries.

Regarding physical features, such as lack of space, location in the northern hemisphere (Figure 2.1), and problems related to depopulation and snowfall, Japan is comparable to some European countries. In order to analyze the compact city model, I choose 3 different areas in Japan, a small, medium and a large scale MtA. The smallest MtA corresponds to Aomori, which operates as a compact city. These 3 MtAs are also located in the northern hemisphere, between the latitude interval of 37.5°N to 45°N and are similar in location as Italy, Spain, Portugal, Northern Greece, Albania, France.

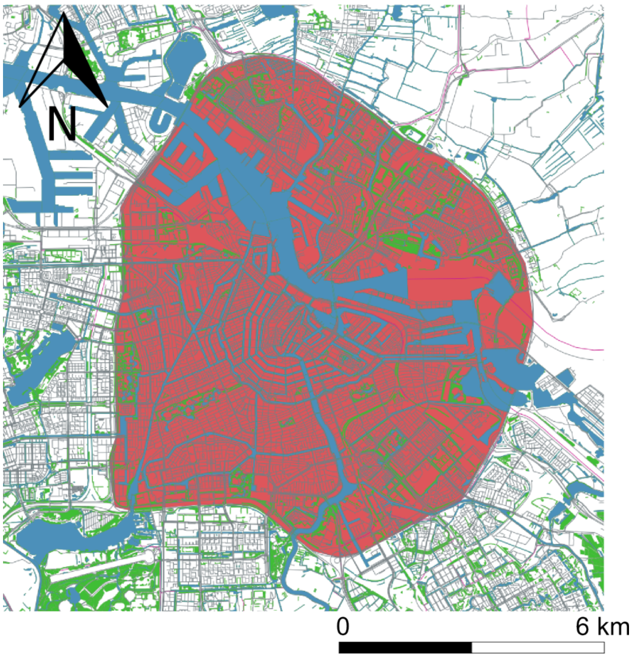


Figure 2.2: Amsterdam an example of European compact city.

European cities such as Oslo in Norway, Groningen, Delft and Amsterdam (Figure 2.2) in the Netherlands, and Halle in Germany, as well as Oxford in U.K., are examples of cities pursuing the compact city model. Multi-functionality is sought through the integration of land uses, transformation to urban mobility, harmonizing of spatial-functional structures and the public transit system [80]. Another example of an European city is Stockholm, Sweden, which seeks to preserve green areas of the islands upon which the city was built and the blue waters surrounding them [67]. From the 1920's the government considered how to improve and preserve some areas for recreational purposes, for that reason decided to purchase those lands. Later, one of the most important investments was to develop an efficient public transport system, based on a network of metro and commuting trains. With this improvement, 70% of people commuting to work from the suburbs to the inner city now use public transportation during rush hours. Another important point worth mentioning is that the city reused and redeveloped its old factories and harbors.

Currently, Japan is faced with important problems related to depopulation and aging, and as a result, it is necessary to gather new residents. Another important problem is related to architecture, because the core of some Japanese cities is still comprised of wooden houses and new buildings, for that reason it is important to improve the city foundations and disaster prevention policies. This problem and the unregulated expansion of the city have led to the loss of farmland and green spaces around the city.

In Japan, some cities have officially incorporated a compact city model, such as Sapporo, Wakkanai, Aomori, Sendai, Toyama, Toyohashi, Kobe and Kitakyushu.

The aim of Japanese compact cities can be divided into five main goals, namely:

1. Addressing the issue of aging
2. Prevention of suburbanization
3. Preservation of city history and culture with special attention to the city core
4. Conservation of nature and environment, and
5. Identification of the current status and future of regional collaboration.

There are some differences between Japanese planners and their counterparts in Europe. For example, Japanese planners do not have the authority to enforce their ideas or projects within the cities- this occurs from the direction and growth at the urban fringe, and constitutes a barrier to planned development of a sustainable urban form [28].

2.2 LEGAL CITY PLANNING SYSTEM IN JAPAN

The legal city planning system in Japan consists of three components: "land use" , "Urban facilities" and "Urban development project"(Ministry of Land, Infrastructure, Transport and Tourism (MLIT)[1].

On the other hand, the urban development project includes "Land Readjustment Project", "Urban Redevelopment Project", "New Residential Area Development Project", "Industrial Park Development Project" and "Residential District Development Project".

1 City planning in Japan: http://www.mlit.go.jp/crd/city/sigaiti/kukaku-e/what/chap1.html

Table 2.1: Overview of Urban Development Projects in Japan (As of March, 1995)

Type	Laws (enacted year)	Implementing Body	Measure	Implementation	
				Number of Projects	Project Area
Land Readjustment Project	Land Readjustment Law (1954)	Local Government, Cooperative, etc.	Replotting	10,254	358,231 ha
Urban Redevelopment Project	Urban Redevelopment Law (1969)	Local Government, Cooperative, etc.	Right Conversion	554	893 ha
New Residential Area Development Project	New Residential Development Law (1963)	Local Government.	Land Acquisition	45	15,591 ha
Industrial Park Development Project	Laws Concerning Development of Suburban and Urban Development Areas in the National Capital Region and the Kinki Region (1958,1964)	Local Government.	Land Acquisition	48	8,308 ha
Residential District Development Project	Special Measures Law for Facilitating Supply of Housing and Residential Lots in Metropolitan Area (1975)	Local Government, Cooperative, etc.	Replotting	7	50 ha

Figure 2.3: Framework of Legal City Planning in Japan (As of March, 1995)

2.3 FRAMEWORK

This proposal of an evaluation method of a compact city model will be focused on two basic conditions. The first is the one related to population distribution. Analyzing the spatial distribution of population by geostatistical methods in different periods of time allows to describe the density of the urban area, and the population behavior nearby the boundaries of the Urbanization Promotion Area (UPA) in each Metropolitan Area (MtA). Also, the spatio-temporal analysis allows to calculate and show the temporal changes. Finally, analyzing population data, the problems related to depopulation and aging will be evident.

In this study, I focus on a metropolitan area. It is defined as a region consisting of a densely populated urban core and its less populated surrounding territories, sharing industry, infrastructure and housing[77]. This region is constituted by "cities, towns and villages"; the surrounding local governments that are essentially urban in character are socially and economically integrated with the central city[63]. I analyze the MtA rather than the cities, because in accordance with the UN there is not a precise definition about a city[2]. Some city planners emphasize different aspects such as geographical, economical, demographic among others. There are also definitions that include a "continuously built up area", or regions closely tied to a central city from the economical point of view. Nevertheless, the previous definitions are similar to the MtA, for that reason I focus on the urban area of the MtA.

2 What is a city? http://www.un.org/cyberschoolbus/habitat/units/un01txt.asp

In other countries the urban planning system is different than the Japanese one. For that reason and in order to make this evaluation method applicable in other countries, it is possible to use Densely Inhabited Districts (DID) instead the UPA. The DID is defined by the government as a district containing basic unit blocks with a population density of 4,000 or more per square kilometer, such districts being adjacent to each other in a municipality[3]. The DID provides the urban area boundaries[26] and it has almost the same size as the urbanization promotion area[47]. While the UPA represents the real city area, the DID represents the common area.

The second basic condition is related to socio-economic factors. One of the goals of the compact city model is the activity concentration. This is linked to the daily facilities such as the ones prepared to provide social, health, meals and therapeutic services, and they should be well distributed in the urban area in order to reduce traveling time. Also, the public transportation intensification such as bus stops is required to reduce the CO_2 emissions and thus avoid traffic congestions. The relationship between the variables taken into account on this study is shown in Figure 2.4.

In chapter 3 the application of the residual kriging model will be explained. Using population and land use information the model and standard error are calculated. To build the kriging model, a variogram is evaluated in order to describe the dissimilarity between observations. Two more indexes will be considered: compactness and entropy. The interpretation of those indexes help to understand the compact city model application and development in the different MtAs.
Chapter 4 highlights the evaluation focusing on socio-economic factors and land use. A classification of land use was calculated by using SVM and later calculating the AUC. A special challenge was addressed in a large MtA, for that reason it was necessary to use parallel computation to deal and solve this problem. Chapter 5 integrates main results from previous chapters to create the evaluation method of a compact city model. Taking into account the results mentioned before, chemical and thermodynamics principles are applied giving a new perspective of the evaluation of the compact city model.
Finally, chapter 6 concludes the present study showing findings through all the evaluation method; future work and suggestions will also be addressed.

3 What is a Densely Inhabited District? http://www.stat.go.jp/english/data/chiri/did/1-1.htm

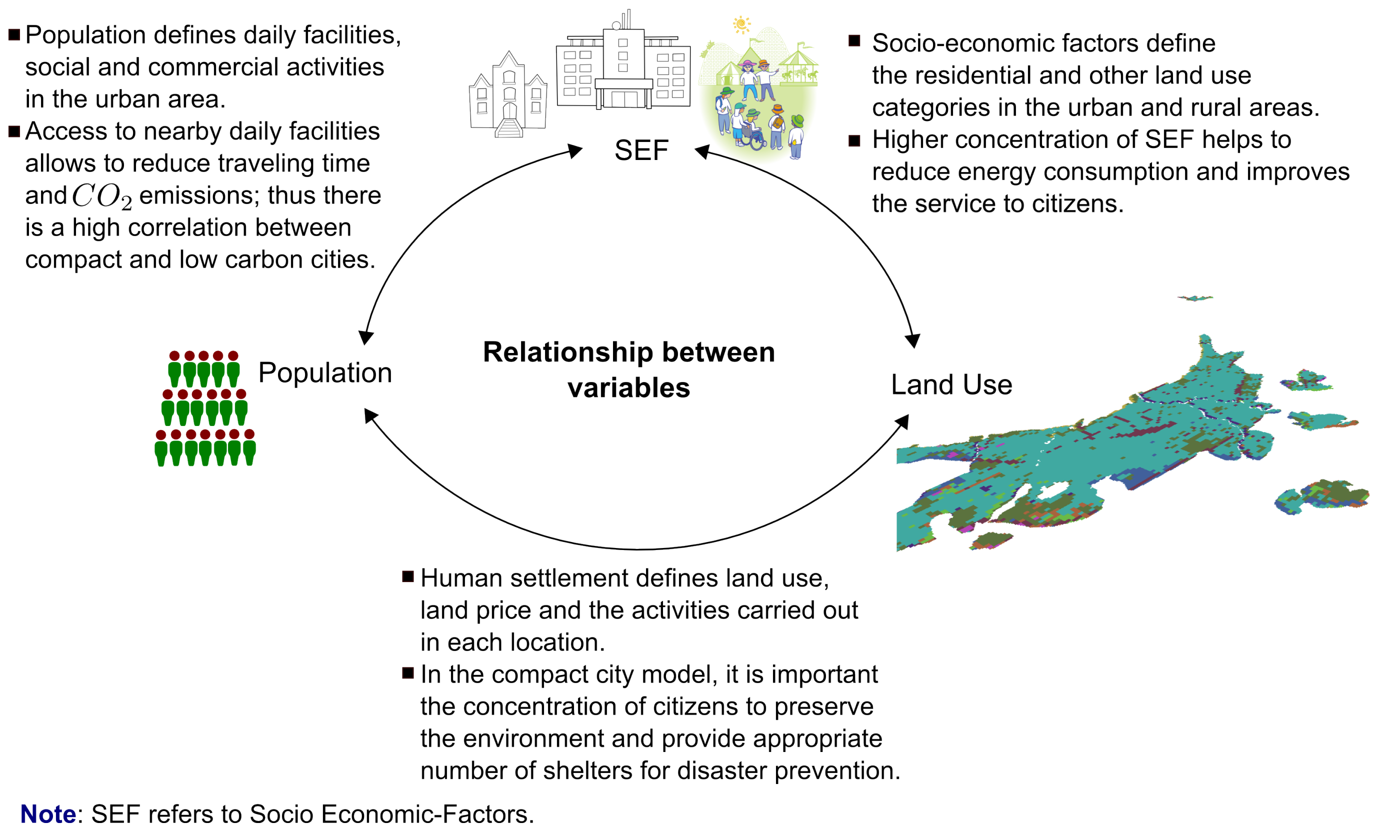


Figure 2.4: Relation between variables

2.4 DESCRIPTION OF STUDY AREA

2.4.1 *Reason for study area selection*

Squires (2002) [86] defines a MtA as a region with high population density in the urban core and a less populated perimeter, with shared industry, infrastructure and housing. The Statistics Bureau of Japan also defines a MtA as one or more central cities which have social cohesion, special wards and ordinance designated cities and their surrounding municipalities. In Japan, there are 14 MtAs, consisting of three major MtAs and other local MtAs. Local MtAs include the Hokkaido, Tohoku, Hiroshima and Fukuoka regions. In Hokkaido, Sapporo is the prefectural capital city and the largest local city in Japan.

The reason for selecting Aomori, Sendai and Sapporo MtAs is that all these areas share common features as such as history, geographical location, problems related to snowfall (e.g. problems on roads) which affects the city budget and problems related to depopulation and aging as life expectancy in Japan reaches almost 83, 10 years longer than in other countries. Aomori and Sendai MtAs are located in the Tohoku area, while Sapporo is on the island of Hokkaido. The difference in the size of the areas is evident - Aomori is equivalent to 6,031 ha (60.31 km^2), while Sendai and Sapporo are 31,718 ha (317.18 km^2) and 74,023 ha (740.23 km^2) respectively. Aomori MtA, which has been working as a compact city since 2000, is a small scale city. Sendai and Sapporo are mid-size and large cities and their related MtAs represent an important section of the Tohoku and Hokkaido regions respectively. Another important aspect is that Sendai and Sapporo MtAs have high population densities and have been studying how to reduce urban sprawl in their areas as a way to improve quality of life.

2.4.2 *Master Plan Study of Metropolitan Areas*

In Japan, each city has its own master plan consisting of a transportation system, community facilities, parks and open spaces, neighborhoods and housing, economic development and land use. According to the Ministry of Internal Affairs and Communications of Japan[4], city planning areas are classified into different areas, such as undivided city planning areas like UPA or Urbanization Control Area (UCA), and divided city planning areas such as District or Zoning areas. An UPA is designated as an industrial, commercial and residential area,

4 Dwelling Environment and Type of City Planning-related terms: http://www.stat.go.jp/english/data/jyutaku/20023.htm

and is defined as an area which already forms an urban area or that will be urbanized within 10 years.

In 2000, a reform of zoning responsibilities was carried out. Today, prefectures may freely decide whether to designate a zone or not. An UCA is designated for agricultural activity and land use is regulated by plans. For a UPA, use of districts is specified for the entire area. However, for the UCA and outside zoning areas, land use districts are only specified for some areas and their management conforms to that of the UPA. Outside zoning areas, which are excluded from the use district, are called white areas [100]. Also, municipalities may designate a quasi-city planning areas[5]; those are the areas that may present an obstacle in the future for the town development. Outside of city planning area is an area that there is no possibility for urban activities; it is regarded as a city of integrated urban planning area that should be specified in advance.

In order to derive the calculations for the city planning areas:

1. Total area = Outside city planning area + outside zoning area + UPA + UCA = 100%.

2. Use district = Outside of zoning area + UPA - white area.

3. Inside city planning area = Outside of zoning area + UCA + UPA

4. The white area = Inside the city planning area - (UCA + Use district).

5. Outside of city planning area = 100% - Inside city planning area.

Figure 2.5 shows the structure of the city planning area.

5 City planning act. Act No. 100 of June 15, 1968

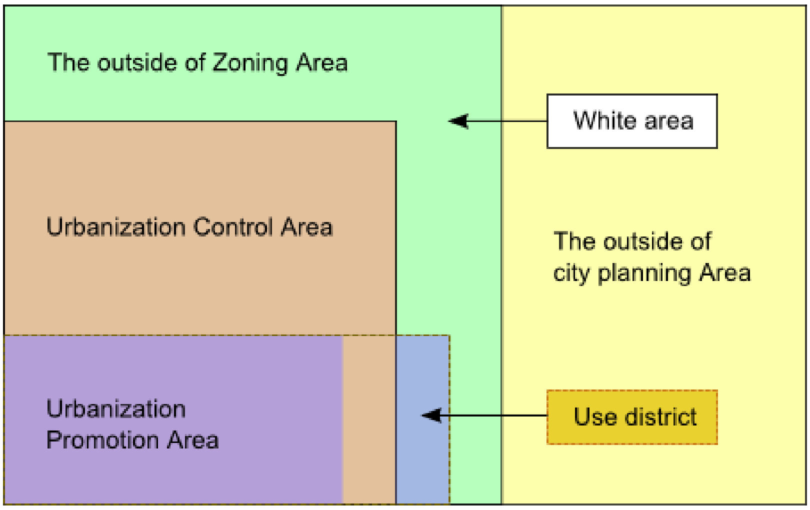


Figure 2.5: City Planning area structure[100]

2.5 DATA

2.5.1 *Population data*

In order to analyze the relationship between the population and land use, it was necessary to get the most detailed information related to population. The Japanese government runs a national census every 5 years. For that reason, and with the aim of keeping as accurate as possible this study, I have taken the information from the periods 1990, 1995 and 2005. This is because the information related to the land use was developed in the years 1991, 1997 and 2006.

The information associated with population in the years 1995 and 2005 is downloadable from the Statistics Bureau that belongs to the Ministry of Internal Affairs and Communications[6]. Information for 1990 was not available for download, but was obtained from the Statistics Bureau of Japan.

Over time and depending on the economic conditions of each region, municipalities are merged or dissolved. It was important to construct the extent of this area in order to determine which cities were merged or became independent. Detailed information for each district was obtained, such as: key code, city name, district name, population by age, latitude, longitude, area, perimeter and density. The information was divided into 4 different groups, as shown in Table 2.2.

Population data in the different periods of time uncovers problems associated with depopulation and aging in each MtA (Table 2.7 and Figure 2.6). The total number of the children and middle age classes in all MtAs have decreased, while the mature and elder classes have

6 Statistics Bureau: http://www.stat.go.jp

Table 2.2: Population group categories

Population group	Range	Code
Children	0-19 years old	CH
Middle age	20-39 years old	MA
Mature people	40-59 years old	MP
Elderly people	60 years old onwards	EP

increased for each year. According to records from the national census, the total population in Aomori MtA has increased, but only by 8.3%, while Sapporo and Sendai have increased growth by 28.7% and 29.2%, respectively, meaning that there was a merging of municipalities.

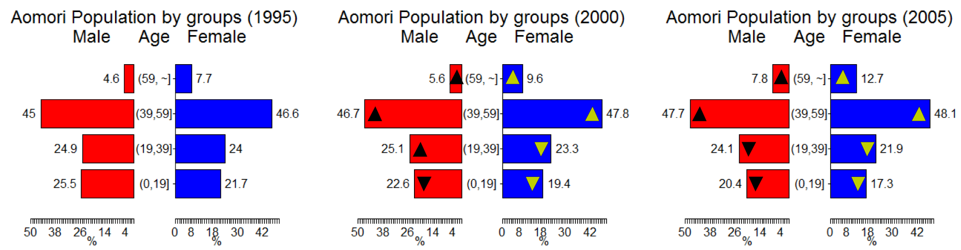
Table 2.4 shows the information given by the population data. However geocoding was needed because the information gives a specific address, but not the information about latitude and longitude. For that reason it was needed to convert the previous information. This data was downloaded from the GIS homepage that belongs to the MLIT[7].

Detailed information related to geo-location is shown in Table 2.5, an example of this data is shown in Table 2.6. For this study the closest years for our targets (land use and population) are taken into account. This data was also downloaded from the GIS homepage from MLIT, in the location reference service[8]. Once the information from the MtAs has been downloaded and unified, it was necessary to merge the data.

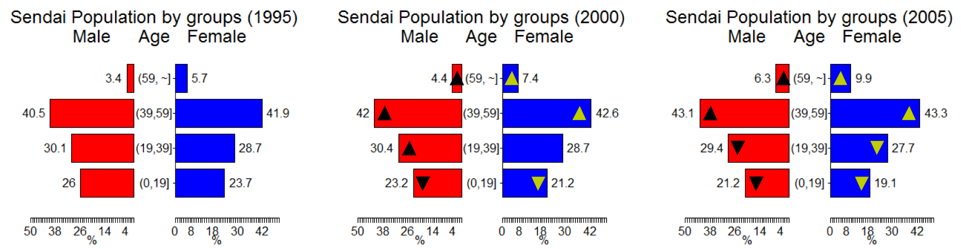
7 GIS homepage: http://www.mlit.go.jp/kokudoseisaku/gis/index.html
8 Location reference information homepage: http://nlftp.mlit.go.jp/isj/index.html

Table 2.3: Population statistics in the MtAs

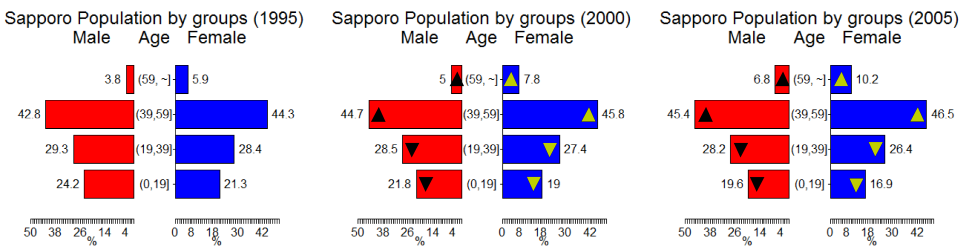
City	Year	Total				Men				Woman				Total records
		CH	MA	MP	EP	CH	MA	MP	EP	CH	MA	MP	EP	
Aomori	1,995	184,531	192,088	360,703	49,284	93,651	91,638	165,394	16,946	90,880	100,450	195,309	32,338	1,671
	2,000	169,054	195,107	382264	62,474	85,782	95,012	176,908	21,277	83,272	100,095	205,356	41,197	1,763
	2,005	149,237	182,589	381,960	83,005	75,579	89,134	176,768	29,056	73,658	93,455	205,192	53,949	1,810
Sendai	1,995	425,325	503,228	705,684	78,578	219,204	254,200	342,149	29,082	206,121	249,028	363,535	49,496	2,510
	2,000	414,316	551,461	789,683	111,019	213,036	278,711	385,134	40,553	201,280	272,750	404,549	70,466	2,841
	2,005	386,308	548,044	829,577	156,339	198,532	275,203	403,883	58,672	187,776	272,841	425,694	97,667	3,243
Sapporo	1,995	590,568	750,429	1,134,651	127,934	302,283	366,307	535,251	47,843	288,285	384,122	599,400	80,091	6,584
	2,000	622,462	854,783	1,384,540	197,288	318,704	416,924	653,997	72,693	303,758	437,859	730,543	124,595	8,508
	2,005	567,262	849,059	1,433,571	267,132	290,609	417,543	673,856	100,804	276,653	431,516	759,715	166,328	8,474



(a) Population pyramid for Aomori by groups (1995, 2000, 2005)



(b) Population pyramid for Sendai by groups (1995, 2000, 2005)



(c) Population pyramid for Sapporo by groups (1995, 2000, 2005)

Figure 2.6: Population pyramids for MtAs

Table 2.4: Information about population

Variable	Type of data
Cities, wards, towns, and villages code	numerical
Basic unit division number	numerical
Investigation division number	alphanumerical
Prefecture name	text
Cities, wards, towns, and villages name	text
Small town name	text
district, neighborhood name	text
Densely-populated-district mark	numerical
Total (according to sex)	numerical
Men	numerical
Women	numerical
Number of households	numerical

Table 2.5: Information about Geo-location

Item	Definition
State code	JIS - State code
State name	Prefectures name of the range
City code	JIS - municipality code
City name	Name of the municipality and range Rural districts including county Name, also district of large town(ku) of ordinance-designated cities)
Small town and neighborhood code	(7 digit JIS municipal district towns and villages code), small town, district and neighborhood code}
Small town, district and neighborhood name	Small town, district and neighborhood, name of the range (Number of the town-district in kanji)
Latitude	Decimal longitude and latitude (unit degree: 6 decimal places, single-byte)
Longitude	Decimal longitude and latitude (unit degree: 6 decimal places, single-byte
Original text material code	Code that represents the original text material in, small towns, district and neighborhoods 1: municipality material, 2: town-ku level position reference information, 3:1 / 25000 terrain Figure, 0: Others material
Small town name, district and neighborhood	Code that represents the distinction of small towns, districts and neighborhoods 1: small town, 2: district, 3: neighborhood, 0: unknown

Table 2.6: Geo-coding example

EXAMPLE OF GEOLOCATION

都道府県コード	都道府県名	市区町村コード	市区町村名	大字町丁目コード	大字町丁目名	緯度	経度	原典資料コード	大字・字・丁目区分コード
1	北海道	1101	札幌市中央区	11010001001	旭ヶ丘一丁目	43.041403	141.31998	1	3
1	北海道	1101	札幌市中央区	11010001002	旭ヶ丘二丁目	43.039804	141.321595	1	3
1	北海道	1101	札幌市中央区	11010001003	旭ヶ丘三丁目	43.039789	141.319717	1	3
1	北海道	1101	札幌市中央区	11010001004	旭ヶ丘四丁目	43.038765	141.3228	1	3
1	北海道	1101	札幌市中央区	11010001005	旭ヶ丘五丁目	43.037356	141.322718	1	3
1	北海道	1101	札幌市中央区	11010001006	旭ヶ丘六丁目	43.037008	141.31897	1	3
1	北海道	1101	札幌市中央区	11010002000	円山	43.048356	141.314493	1	1
1	北海道	1101	札幌市中央区	11010003001	円山西町一丁目	43.048663	141.303129	1	3
1	北海道	1101	札幌市中央区	11010003002	円山西町二丁目	43.044979	141.304846	1	3
1	北海道	1101	札幌市中央区	11010003003	円山西町三丁目	43.043095	141.305022	1	3
1	北海道	1101	札幌市中央区	11010003004	円山西町四丁目	43.040237	141.304668	1	3
1	北海道	1101	札幌市中央区	11010003005	円山西町五丁目	43.038045	141.304161	1	3
1	北海道	1101	札幌市中央区	11010003006	円山西町六丁目	43.041294	141.307533	1	3
1	北海道	1101	札幌市中央区	11010003007	円山西町七丁目	43.042746	141.309327	1	3
1	北海道	1101	札幌市中央区	11010003008	円山西町八丁目	43.04504	141.306831	1	3
1	北海道	1101	札幌市中央区	11010003009	円山西町九丁目	43.044365	141.311477	1	3
1	北海道	1101	札幌市中央区	11010003010	円山西町十丁目	43.046419	141.30967	1	3
1	北海道	1101	札幌市中央区	11010004000	円山西町	43.041677	141.300397	1	1
1	北海道	1101	札幌市中央区	11010005001	界川一丁目	43.041596	141.318196	1	3
1	北海道	1101	札幌市中央区	11010005002	界川二丁目	43.035306	141.316405	1	3
1	北海道	1101	札幌市中央区	11010005003	界川三丁目	43.036151	141.314385	1	3
1	北海道	1101	札幌市中央区	11010005004	界川四丁目	43.039749	141.313349	1	3
1	北海道	1101	札幌市中央区	11010006001	宮ケ丘一丁目	43.058257	141.312369	1	3
1	北海道	1101	札幌市中央区	11010006002	宮ケ丘二丁目	43.057968	141.310945	1	3
1	北海道	1101	札幌市中央区	11010006003	宮ケ丘三丁目	43.05761	141.309369	1	3
1	北海道	1101	札幌市中央区	11010007000	宮ケ丘	43.053568	141.308622	1	1
1	北海道	1101	札幌市中央区	11010008000	宮の森	43.050177	141.287485	1	1
1	北海道	1101	札幌市中央区	11010009001	宮の森一条一丁目	43.06457	141.312921	1	3
1	北海道	1101	札幌市中央区	11010009002	宮の森一条二丁目	43.063628	141.312063	2	3
1	北海道	1101	札幌市中央区	11010009003	宮の森一条三丁目	43.062833	141.311264	2	3
1	北海道	1101	札幌市中央区	11010009004	宮の森一条四丁目	43.06223	141.310745	2	3
1	北海道	1101	札幌市中央区	11010009005	宮の森一条五丁目	43.061658	141.310427	2	3
1	北海道	1101	札幌市中央区	11010009006	宮の森一条六丁目	43.060622	141.309138	1	3
1	北海道	1101	札幌市中央区	11010009007	宮の森一条七丁目	43.059536	141.309176	2	3
1	北海道	1101	札幌市中央区	11010009008	宮の森一条八丁目	43.058889	141.30853	2	3
1	北海道	1101	札幌市中央区	11010009009	宮の森一条九丁目	43.058204	141.307707	2	3
1	北海道	1101	札幌市中央区	11010009010	宮の森一条十丁目	43.056954	141.306136	1	3
1	北海道	1101	札幌市中央区	11010009011	宮の森一条十一丁目	43.056276	141.303997	2	3
1	北海道	1101	札幌市中央区	11010009012	宮の森一条十二丁目	43.054302	141.302979	2	3
1	北海道	1101	札幌市中央区	11010009013	宮の森一条十三丁目	43.052432	141.302162	1	3
1	北海道	1101	札幌市中央区	11010009014	宮の森一条十四丁目	43.050705	141.301822	1	3
1	北海道	1101	札幌市中央区	11010009015	宮の森一条十五丁目	43.048639	141.297281	1	3
1	北海道	1101	札幌市中央区	11010009016	宮の森一条十六丁目	43.046518	141.295128	1	3
1	北海道	1101	札幌市中央区	11010009017	宮の森一条十七丁目	43.043654	141.292658	1	3
1	北海道	1101	札幌市中央区	11010009018	宮の森一条十八丁目	43.040966	141.291464	1	3
1	北海道	1101	札幌市中央区	11010010001	宮の森三条一丁目	43.066669	141.309263	2	3
1	北海道	1101	札幌市中央区	11010010002	宮の森三条二丁目	43.065734	141.308901	2	3
1	北海道	1101	札幌市中央区	11010010003	宮の森三条三丁目	43.064873	141.307563	2	3
1	北海道	1101	札幌市中央区	11010010004	宮の森三条四丁目	43.064482	141.3068	2	3
1	北海道	1101	札幌市中央区	11010010005	宮の森三条五丁目	43.063594	141.306219	2	3
1	北海道	1101	札幌市中央区	11010010006	宮の森三条六丁目	43.063155	141.305309	2	3
1	北海道	1101	札幌市中央区	11010010007	宮の森三条七丁目	43.061893	141.304058	2	3
1	北海道	1101	札幌市中央区	11010010008	宮の森三条八丁目	43.061278	141.302712	2	3
1	北海道	1101	札幌市中央区	11010010009	宮の森三条九丁目	43.060574	141.301964	2	3

Table 2.7: Land use categories

Code	Description	Color
1	Rice fields	
2	Other agricultural land	
3	Forest	
4	Waste Land	
5	Buildings,residential areas	
6	Roads	
7	Other sites	
8	Rivers and lake areas	
9	Beach	
10	Ocean	
11	Golf course	

2.5.2 *Land use data*

The information used for this study is related to the land use master plan of the different MtAs, and it was collected from the Land-Use Control Back-up System (LUCKY) provided by MLIT. This system classifies different types of land use in the Japanese geography, and also it is possible to extract the UPA. In order to reduce calculation errors, it was necessary to analyze the image files (raster files) with an automated process [76] using GIS. These processes are regularly used in remote sensing, and allow the extraction of specific features.

Land use classification data was downloaded from the National Land Numerical Information download service which is also provided by MLIT for 2006. It is measured by 100 x 100m grid, and the classification system has a unique value per mesh area. The central government in Japan has defined land use into 11 different categories (Table 2.7), excluding mixed type land use. For the present analysis, respective codes and colors for each type of land use are used.

Using GIS, I overlaid the UPA's shapefile with the land use data to extract the land use information in the UPA. Figures (2.7, 2.8, 2.9) present the UPA of Aomori MtA; Figures (2.10, 2.11, 2.12) present the UPA of Sendai MtA and Figures (2.13, 2.14, 2.15) present the UPA of Sapporo MtA in different periods of time.

In Table 2.8, the percentage corresponding to the UPA of Aomori MtA is clearly the smallest among the three MtAs, however also has the highest percentage area outside of city planning. This is because Aomori Prefectural government wants to keep the UPA just for urban development projects. It is clear that Sendai MtA has the highest percentage area inside the city planning area and the areas of Aomori and Sapporo MtAs are similar.

Table 2.9 presents the statistics about land use, it is evident that building and residential area occupies more than 63% of the total area, while the area designated to other types of land occupies just 15%. In Table 2.10, land use changes for each UPA are shown for the years 1991, 1997 and 2006. The largest change related to rice fields was in Aomori MtA, with a reduction greater than 55%, but residential areas and roads also grew more than 25% and 23%, respectively. However, there was a reduction in agricultural fields and Sendai MtA experienced the largest change at more than 67% as well as a large growth in residential areas at more than 70%. It is evident that areas such as paddy fields and those for other agricultural purposes decreased while residential areas increased.

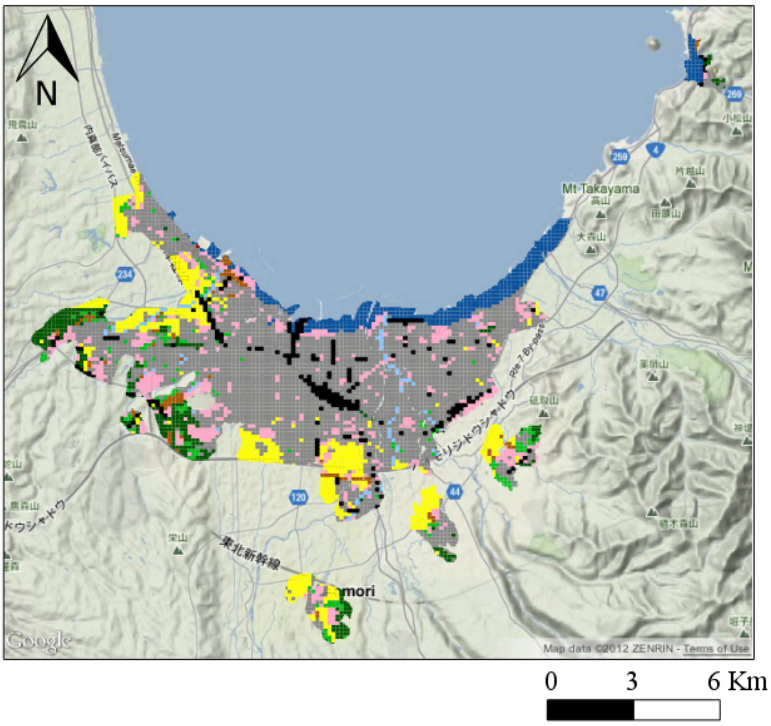


Figure 2.7: UPA of Aomori MtA 1991

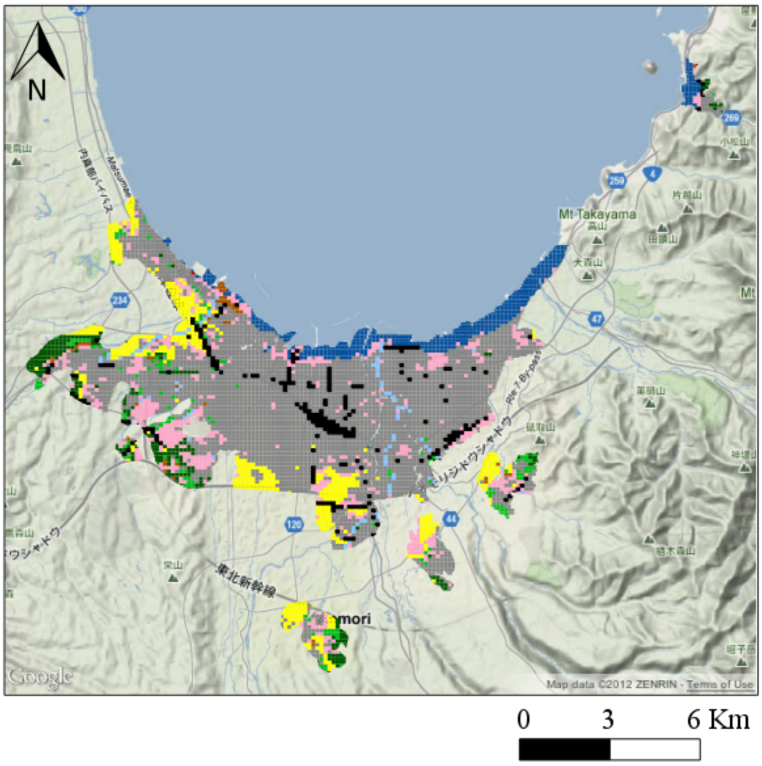


Figure 2.8: UPA of Aomori MtA 1997

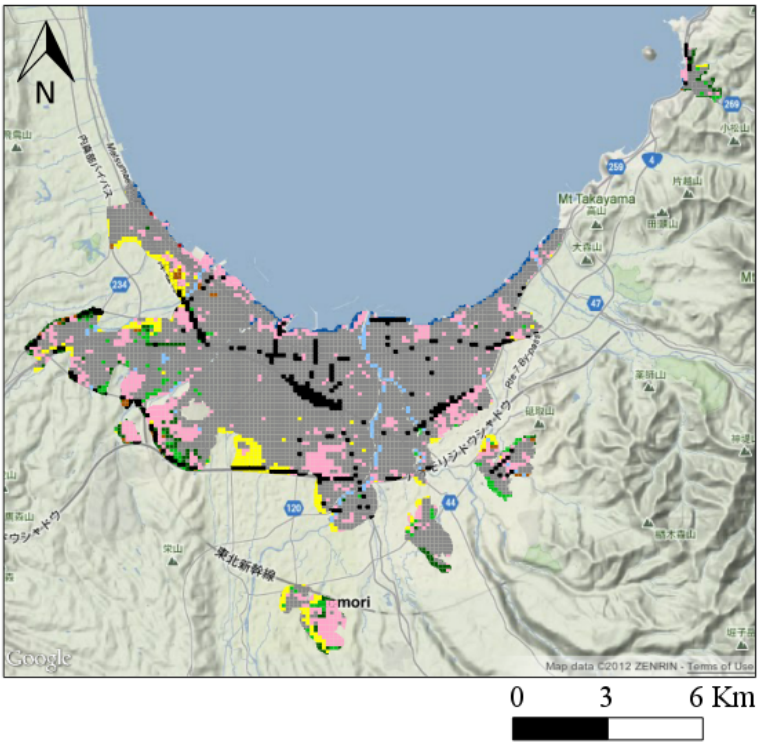


Figure 2.9: UPA of Aomori MtA 2006

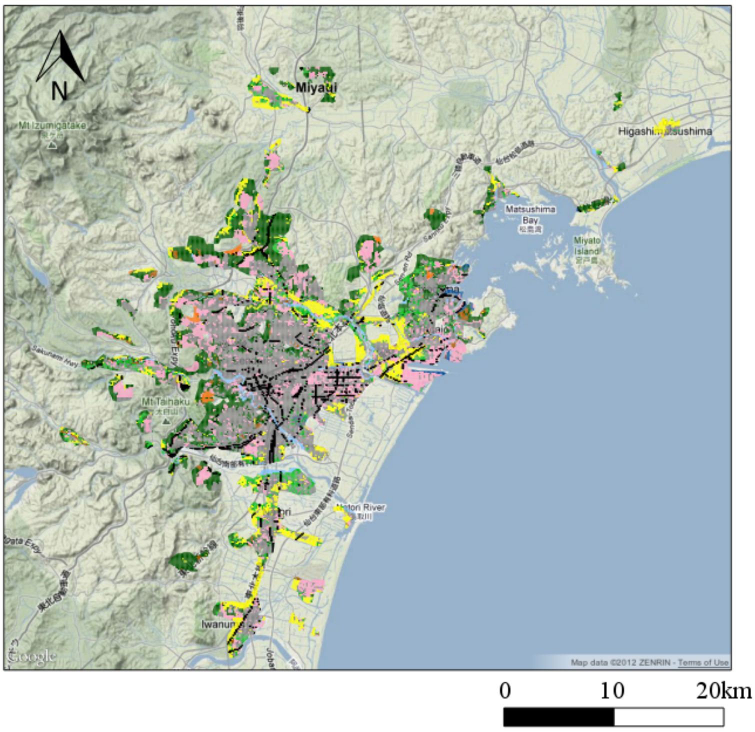


Figure 2.10: UPA of Sendai MtA 1991

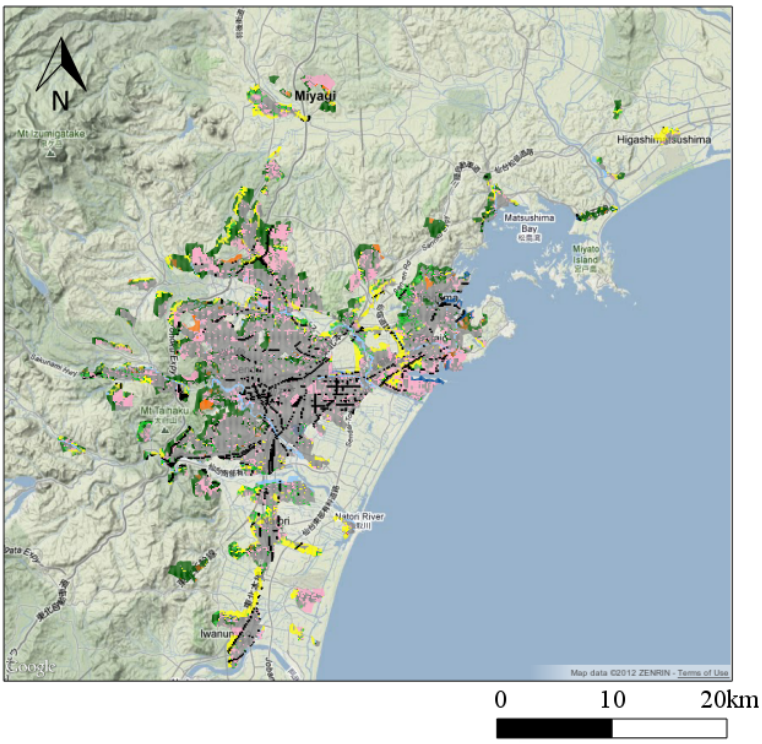


Figure 2.11: UPA of Sendai MtA 1997

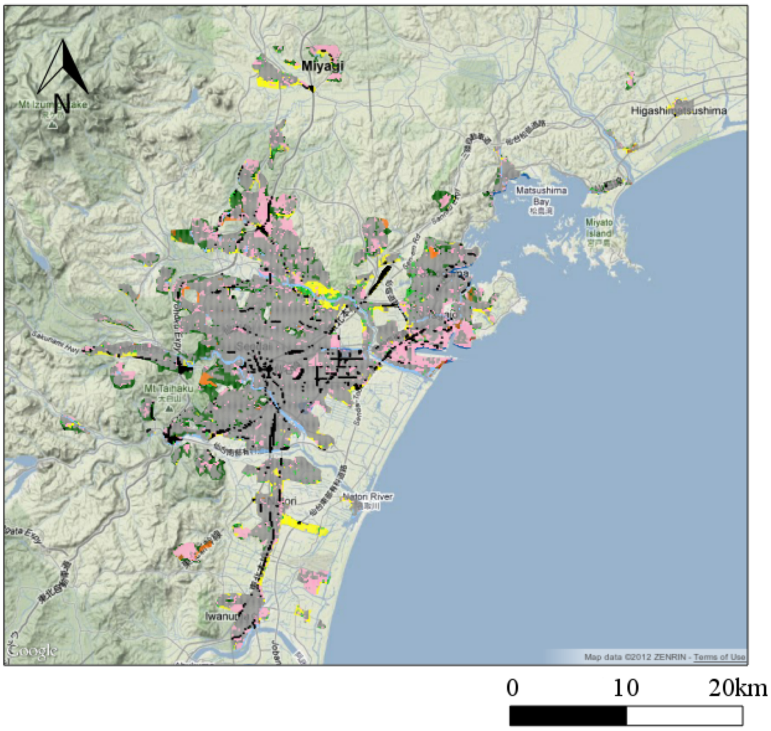


Figure 2.12: UPA of Sendai MtA 2006

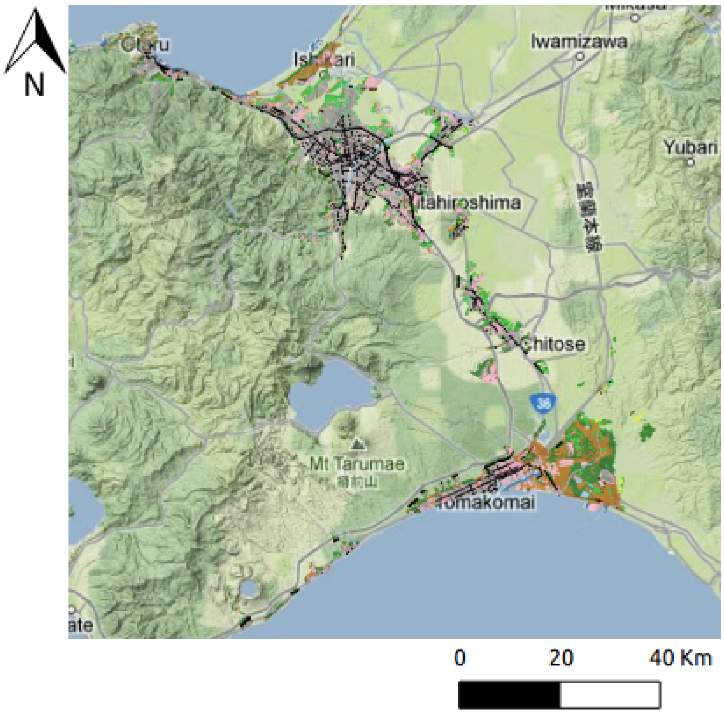


Figure 2.13: UPA of Sapporo MtA 1991

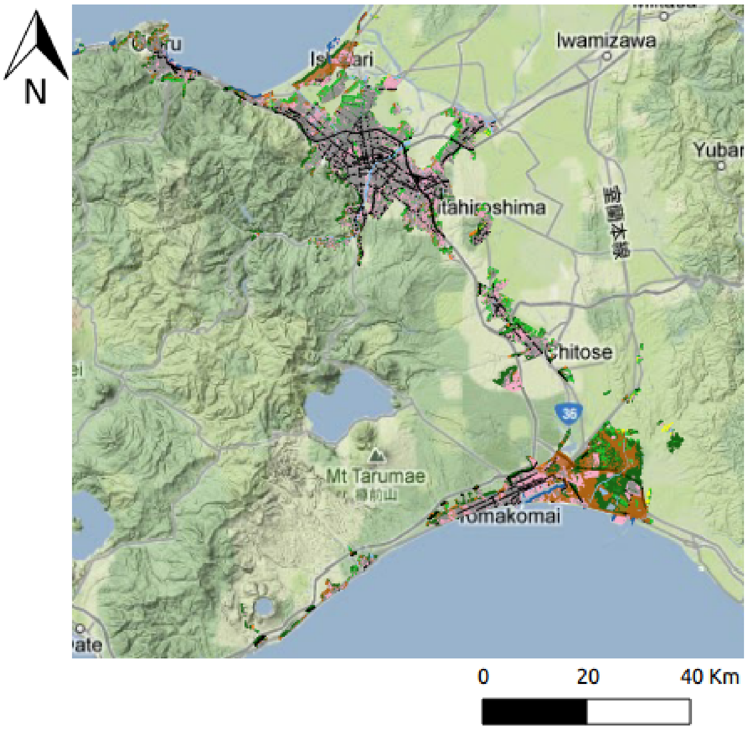


Figure 2.14: UPA of Sapporo MtA 1997

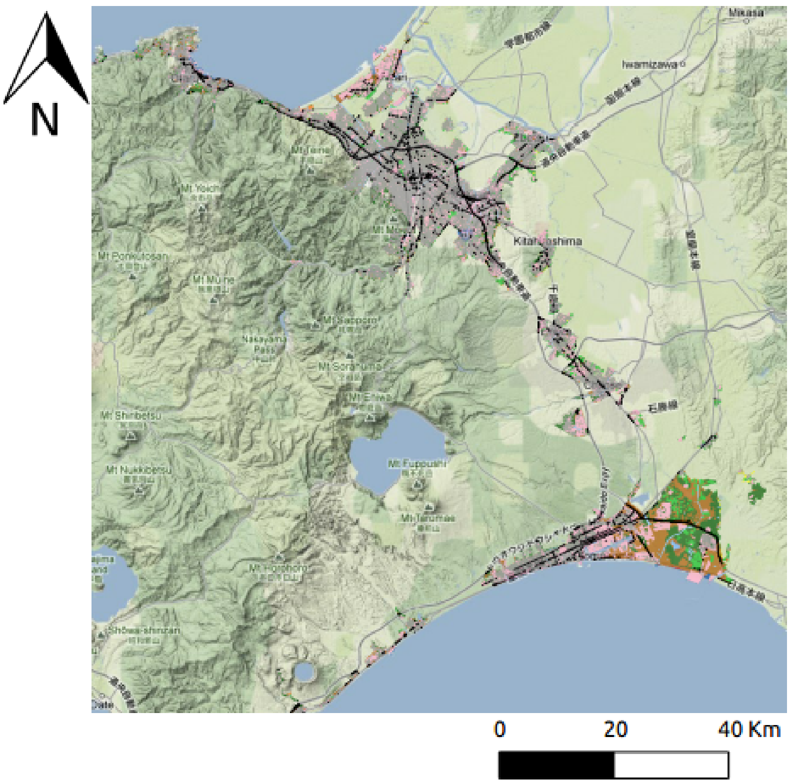


Figure 2.15: UPA of Sapporo MtA 2006

Table 2.8: Percentage of areas with special land use controls for the entire MtA(%)

1	City planning area	Aomori	Sendai	Sapporo
2	UPA	1.4	9.6	6.1
3	UCA	11.2	27.9	18.8
4	Use District	0.9	3	1.4
5	Outside of zoning area	12.9	12.1	3.3
6	Inside city planning area	25.5	49.6	28.2
7	White area	13.3	18.7	8.1
8	Outside city planning area	74.5	50.4	71.8

Table 2.9: Land use in the Urbanization Promotion Areas (ha)

Met. Area	Period	Rice field	Other agricultural	Forest	Waste Land	Building site	Arterial traffic	Other	Rivers lakes	Beach	Ocean	Golf course
Aomori	1991	773	194	374	103	3,068	247	662	90	1	519	0
	1997	655	175	309	51	3,327	245	673	90	1	505	0
	2006	345	140	206	45	3,848	305	889	105	2	146	0
Sapporo	1991	615	7,974	12,055	10,789	21,696	2,543	15,020	1,893	83	1,079	276
	1997	430	6,113	10,534	7,863	28,982	3,017	13,797	1,858	82	1,065	282
	2006	291	3,785	7,896	6,925	35,642	2,849	14,003	1,666	156	714	96
Sendai	1991	4,174	1,824	7,003	430	10,696	1,191	5,177	834	1	173	215
	1997	2,964	1,193	5,231	259	14,802	1,234	4,968	848	264	201	0
	2006	1,954	609	3,301	200	18,517	1,160	4,522	996	22	197	240

Table 2.10: Land use changes in the Urbanization Promotion Areas (ha)

MtArea	Period	Rice field	Other agricultural	Forest	Waste Land	Building.site	Arterial traffic	Other	Rivers lakes	Beach	Ocean	Golf course
Aomori	97-91	-118	-19	-65	-52	259	-2	11	0	0	-14	0
	06-97	-310	-35	-103	-6	521	60	216	15	1	-359	0
sapporo	97-91	-185	-1,861	-1,521	-2,926	7,286	474	-1,223	-35	-1	-14	6
	06-97	-139	-2,328	-2,638	-938	6,660	-168	206	-192	74	-351	-186
Sendai	97-91	-1,210	-631	-1,772	-171	4,106	43	-209	14	263	28	-215
	06-97	-1,010	-584	-1,930	-59	3,715	-74	-446	148	-242	-4	240

2.5.3 *Socio-economic factors*

The studies related with socio-economic factors in Japan are focused on health, diet and mortality. Fukuda et al. [24] studied the sex-specific mortality of municipalities by age groups. They linked this problem with municipal socio-economic status indicators related to income, education, unemployment and living space. Their results showed that the mortality gradient had a higher impact on citizens less than 75 years old population than the total and over 75 years old population, and the relationship between mortality and income-education related indicator was stronger for males than for females. The above mentioned authors continue studying the wide range of socio-economic factors associated with mortality, focusing on factors such as unemployment, old housing, primary health resources and density. They results showed that for women's mortality, higher income, unemployment spacious dwelling, old housing, less vegetation, road facility number of cars per population, primary health resources and density were positively associated. Whereas higher education, public library activity and health check-up participation were independently negatively associated [25].

Other important socio-economic factors that cities have to deal with are: for instance, transportation and road systems, dwelling, industrial contamination of rivers, lakes or coastal zones, degradation of the landscape, shortage of green spaces and public recreation areas and lack of education, training or effective institutional cooperation in environmental management [59]. In accordance with the socio-economic factors before mentioned, I will focus on transportation system such as bus stops and railroad stations, medical institutions, public facilities and I will also study land use price and other important facilities for the Japanese life style.

Socio-economic factors characterize the individual or group within the social structure. Among the most important socio-economic factors are education, income and occupation, place of residence, culture and ethnicity and religion. In this study, socio-economic factors that affect housing decision making were selected. These factors affect and define the activity in each grid area. The data related to the socio-economic factors such as railway stations, bus stops, convenience stores, malls, medical institutions, governmental and public services and land price by district was downloaded from the MLIT and public sources. Detailed information about them is shown in Table 2.11, types *num* and *cat* refer to numeric and categorical data. The price of land is measured by district, however it was necessary to provide the price by 100m mesh areas.

Table 2.11: Socio-economic factors in the different prefectures

Variable	Additional information	Aomori	Miyagi	Hokkaido	Type
Bus stops	None	4,555	2,793	4,099	Num
Latitude, longitude	None	6,030	31,718	74,023	Num
Convenience stores	Circle K Sunkus, Sunkus, Ministop, Lawson	399	287	447	Num
Medical institutions	Ministry of health, labour and welfare, national hospital organizations, Social insurance, pension, sailors insurance, Hospital, Clinic, dental clinic, private medical institutions, welfare corporations	1,522	1,908	3,652	Num
Parks	None	754	2,201	4,240	Num
Land use price	By district	325	531	929	Num
Public Facilities	Amusement parks, Other, National institutions, Local government, Welfare agency, Police agency, Fire station, School, Hospital, Post office, Welfare facilities	3,216	2,202	3,556	Num, Cat
Supermarkets	None	85	42	143	Num
Train stations	None	9	98	179	Num
Land use data	Urbanization Promotion Area	6,031	31,718	74,023	Cat

Detailed information about socio-economic factors in the UPA of Aomori MtA is provided in appendix A.1. The information of socio-economic factors in the UPA is shown in table 2.12.

where:

bust	Bus stops
convst	Convenience stores
medinst	Medical institutions
pubfl	Public facilities
smkt	Supermarkets
trainst	Train stations

Table 2.12: Socio-economic factors in the UPAs

	bust	convst	medinst	others	parks	price	pubfl	smkt	trainst
Aomori	433	83	382	72	120	89	504	12	12
Sendai	2,793	287	1,908	0	2,201	531	2,202	42	98
Sapporo	4,099	447	3,652	0	4,240	929	3,556	143	179

2.5.4 *Data processing*

1. First of all, it was necessary to analyze image files (raster files) with an automated process [76] in a Geographical Information System (GIS) to reduce calculation errors. This process was required to extract the information related with Urbanization Promotion Area (UPA), and Urbanization Control Area (UCA) from the Land Use Backup System (LUCKY).

2. Second, the population data was unified since the national government has changed the format after 1995 census.

3. Third, once the population has been unified, it was overlaid with the geo-coding information in order to create a geospatial vector file. Although there are different systems to calculate the geocoding (Google, Geocoding service provided by Tokyo University[9], Yahoo, among others), in this study the information provided by the National Government was used.

4. Population data was rescaled in order to reduce calculation errors and the density was rescaled and recentered to have a mean of 0 and a variance of 1[49] [60].

5. Using land use data a residual Kriging model for population density was developed, also the standard errors were calculated to identify the location where the model presents more deviation. Depending on the size of the map and grid size, computational time may increase. In the same way as in the first phase, it was necessary to efficiently handle the database in order to reduce time complexity O(N4)[87] or improve the kriging model.

6. As well as population, the socio-economic data was gathered from public and private sources. In order to unify all the data, the socio-economic factors data was overlaid with geo-coding information and converted into geospatial vector file.

Figure 2.16 shows the flow diagram for data processing. On the left side the flow diagram for population data is shown, while the flow diagram for socio-economic data is shown on the right side.

9 CSV address matching service: `http://newspat.csis.u-tokyo.ac.jp/geocode-cgi/geocode.cgi?action=start`, last access: 18/07/2014

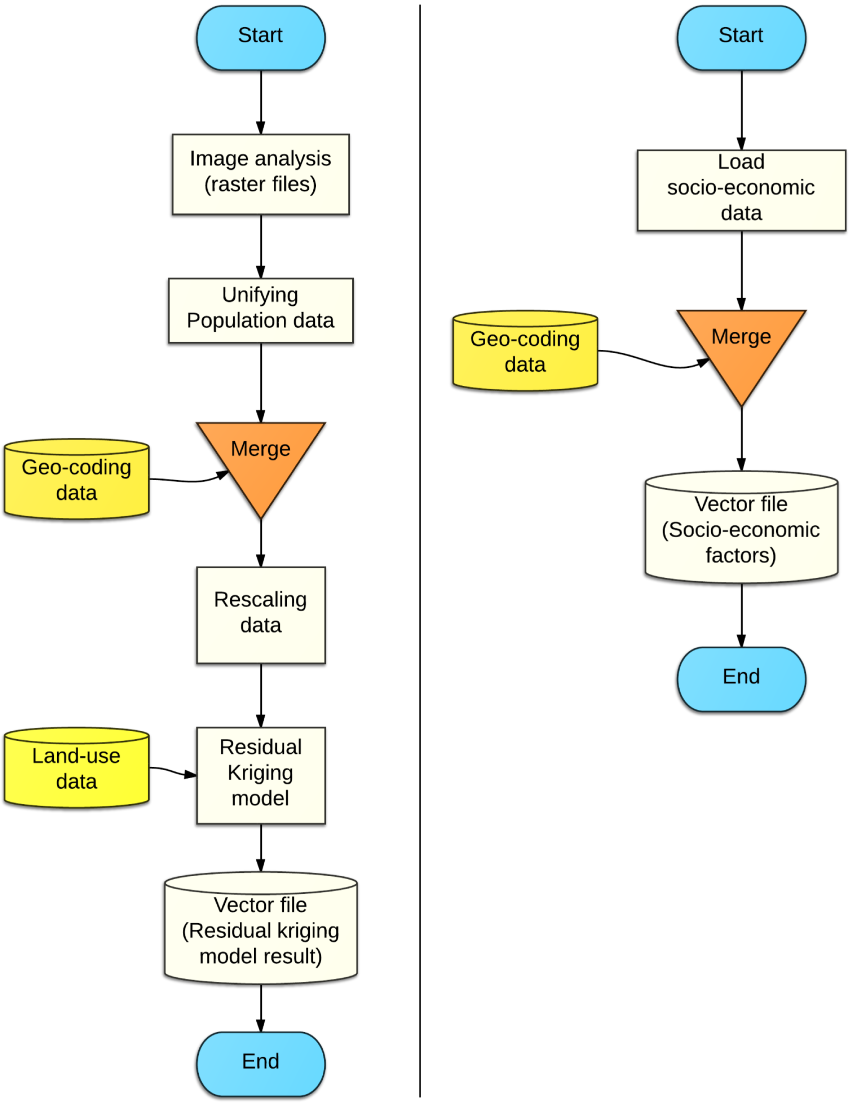


Figure 2.16: Data processing (Flow diagram).

2.6 SOFTWARE

In order to merge and manipulate the database I have used two software solutions:

R[72] is a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc.[10].
The following packages were used to develop the different models: To develop the Kriging model *gstat*[66] package, to run the parallel

10 What is R: http://cran.r-project.org, last access: 02/07/2014

processes *doParallel*, *foreach* were used. The *kernlab*[43] package was used to calculate the SVM model. This tool is useful for kernel-based machine learning methods for classification, regression, clustering, novelty detection, quantile regression and dimensionality reduction. It also includes SVM, Spectral clustering, Kernel PCA. *Caret* package was also used to calculate the variable importance of the predictors, this package is also useful for data splitting, pre-processing, feature selection and model tuning by resampling. The *spatstat* package was useful for analyzing spatial data, mainly spatial point patterns, including multitype/marked points and spatial covariates, in any two-dimensional spatial region[6]. It also supports three-dimensional point patterns, and space-time point patterns in any number of dimensions [7].

ArcGIS 10.1 is a Geographic Information System developed by ESRI corporation, useful to deal with geographical information and digital maps. It is a system to collect, organize, manage, communicate and distribute geographical information[11]. Arcmap is a software to display geographic information in different layers and also other elements on a map. As basic features of this professional library are: working and printing maps, compile and edit GIS datasets, use geoprocessing to automate work and perform analysis, publish maps documents as map services using ArcGIS server, share maps, layers, geoprocessing models and geodatabases[12].

11 What is ArcGIS: http://resources.arcgis.com/en/help/getting-started/articles/026n00000014000000.htm, last access: 02/07/2014

12 What is ArcMap: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//006600000001000000, last access: 02/07/2014

3 EVALUATION FOCUSING ON POPULATION AND LAND USE

This proposal of an evaluation method will be focused on two basic characteristics of the compact city model. In this chapter I will focus on the one related to population distribution. Analyzing the spatial distribution by geostatistical analysis of population in different periods of time allows to describe the density of the urban area, and the population behavior nearby the boundaries of the UPA in each MtA. Also the spatio-temporal analysis allows to calculate and show the temporal changes. Finally, analyzing population data, the problems related to depopulation and aging will be evident.

In this chapter, the relationship between population and land use will be clarified. First of all, it will explain the methodology such as the residual Kriging model, and how all the indexes were calculated. The next part is the explanation about the Compactness index, it will analyze the interaction between the land use and compactness. Finally, a concept originally from thermodynamics and later introduced in information theory will be applied, it is named as Shannon's entropy. This indicator is useful to measure the urban sprawl, an opposite concept of compactness. For that reason, I want to calculate the urban sprawling through Shannon's entropy for each MtA in the different periods of time.

In this chapter, I will inspect the compact city model in the different metropolitan areas with residual Kriging model. Also the compactness and entropy indexes will be evaluated in the different periods of time. It is a necessary step to give a preliminary approach to the compact city model in these metropolitan areas.

3.1 METHODOLOGY

3.1.1 *Variograms*

In order to develop the kriging model it is necessary to describe the population distribution using semi-variograms [30]. A variogram is a function to describe the dissimilarity between observations, and it models spatial correlation and plots semi-variance as a function of distance [8]. It is defined as:

$$\gamma(h) = \frac{1}{2}\mathbb{E}\left(Z(s) - Z(s+h)\right)^2 \tag{1}$$

Where h is the separation distance between pairs of points.

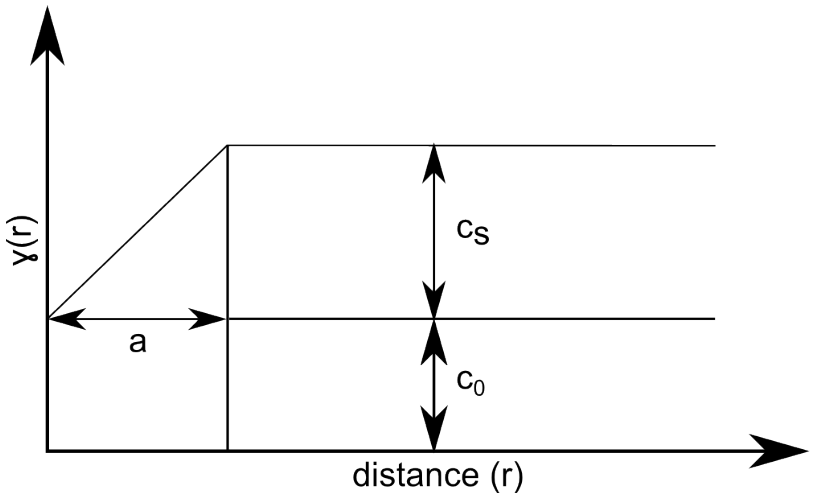


Figure 3.1: Sketch of the parametric variogram and its parameters

According to this, variograms only depend on the separation distance. A sampling of variograms are necessary to evaluate a data sample, these can be estimated through N_h sample data pairs $z(s_i)$, $z(s_{i+h})$ for the number of distances $\tilde{h}$.

$$\hat{\gamma}(\tilde{h}) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (Z(s_i) - Z(s_i + h))^2, \qquad \forall h \in \tilde{h}_j \tag{2}$$

Sample variograms can be estimated using Equation (2).

A parametric semivariogram (Figure 3.1) has 3 parameters: A parameter c_0 represents the "nugget effect", it represents a micro scale variation; c_s is the "partial sill", it is when the semivariogram model attains at the range; finally the variance of random process and a is the "range", the longest distance with correlated values of the random process [18].

3.1.2 *Residual Kriging model*

The Kriging interpolation is a method which predicts unknown values of observed data. $Z(s)$ is defined as a random field, where Z is a random value and s is the non-random spatial index.

$$\widehat{Z}(s_0) = x(s_0)\widehat{\beta} + v'V^{-1}(Z(s) + X\widehat{\beta}) \tag{3}$$

Where X is the matrix of observations and $v'V^{-1}$ are the weights of the model, also known as simple Kriging weights. Equation (3) is the best linear unbiased predictor of $Z(s_0)$. The prediction error variance is defined as follows:

$$\sigma^2(s_0) = \sigma_0^2 - v'V^{-1}v + \delta(X'V^{-1}X)^{-1}\delta' \tag{4}$$

where $\sigma^2(s_0)$ is the variance of $Z(s_0)$, and $\delta = x(s_0) - \nu' V^{-1} X$.

Equation (4) can be written as a function of the lag vector h, in this way the expected value of the difference between $Z(s)$ and $Z(s+h)$ is 0.

$$
\begin{aligned}
Var\{[Z(s+h)+Z(s)]\} &= E\left\{[Z(s+h)-Z(s)]^2\right\} \\
&= 2\gamma_R
\end{aligned}
\tag{5}
$$

Where $2\gamma_R(h)$ is the semi-variogram or residuals [50]. It is necessary to check if the land use zone was referred to its centroid u in order to interpolate the population density usage area to point by residual kriging. The population density is defined as:

$$
d(u) = \widehat{d}(u) + r(u) \tag{6}
$$

Where d(u) is the unknown population density of a land use zone, $\widehat{d}(u) = m(u)$ is the value estimated by the regression model, and $r(u)$ is the residual population density. Let ν_u be the census unit containing u which has a residual density $r(u_\nu)$ that can be defined as:

$$
\begin{aligned}
r(\nu_u) &= \frac{P_{\nu_u} - \sum_j \widehat{d}(u_j) A_{u_j}}{A_{\nu_j}} \\
&= \frac{\sum_j d(u_j) A_{u_j} - \sum_j \widehat{d}(u_j) A_{u_j}}{A_{\nu_u}} \\
&= \sum_j \frac{\left\{ d(u_j) \widehat{d}(u_j) \right\} A_{\nu_u}}{A_{\nu_u}} \\
&= \sum_j \frac{A_{u_j}}{A_{\nu_u}} r(u_j)
\end{aligned}
\tag{7}
$$

Where A_{u_j} is the area of the land-use zone u_j, and P_{ν_u} and A_{ν_u} are the population and area of the host census unit, respectively. The value $\widehat{d}(u_j)$ is calculated from the regression model.

In order to create the model to predict population density in different mesh areas, a functional regression model is estimated to integrate functional response and scalar covariates [78]. The model is defined as follows

$$
Z_{s_i}(t) = \alpha(t) + \beta_1(t) x_i + \beta_2(t) y_i + \epsilon_i(t) \tag{8}
$$

Where $Z_{s_i}, i = 1, \ldots, n$ are the functions at sampled data, (x_i, y_i) are the geographical coordinates, and $\alpha(t), \beta_1(t), \beta_2(t)$ are functional parameters of interest and $\epsilon_i(t)$ refers to noise for each $t \in T$.

However in order to estimate the residual data it is required to calculate the noise in the following way.

$$\epsilon_{s_i}(t) = Z_{s_i}(t) - \widehat{Z}_{s_i}(t) \tag{9}$$

$$\epsilon_{s_i}(t) = Z_{s_i}(t) - \left(\widehat{\alpha}(t) + \widehat{\beta}_1(t)x_i + \widehat{\beta}_2(t)y_i\right) \tag{10}$$

Where $\epsilon_{s_i}(t), i = 1, \ldots, n$ are the residual curves from the functional regression model. Finally, so as to predict a non-sampled location it is required to calculate the following equation

$$\widehat{X}_{s_0}(t) = \widehat{Z}_0(t) + \widehat{\epsilon}_0(t) \tag{11}$$

where:

$\widehat{X}_{s_0}(t)$	The predicted value on the location s_0
$\widehat{Z}_{s_0}(t) = \widehat{\alpha}(t) + \widehat{\beta}_1(t)x_0 + \widehat{\beta}_2(t)y_0$	The trend estimated on the location (x_0, y_0).
$\widehat{\epsilon}_0(t)$	Prediction of a residual function on a non-sampled location.

3.1.3 *Compactness*

In order to calculate compactness for each MtA in the different periods of time, I will calculate the index using Burton's metrics. For this analysis I calculate some relationships before the compactness index.

$$\mathrm{densblt} = \frac{RP_i}{BU} \tag{12}$$

Where RP_i is defined as residential population in zone and BU is the built-up acreage.

$$\mathrm{densres} = \frac{RP_i}{RA} \tag{13}$$

Where RA is the residential area acreage.

$$\mathrm{supfacs} = \frac{RA}{BU - RA} \tag{14}$$

Where equation (14) is the ratio between residential / nonresidential urban land. Finally, the index of compactness is calculated by the ratio between the overall density (OD) and mix-of-use measures (MixU), shown in equation 15.

$$\mathrm{Compactness} = \frac{OD}{MixU} \tag{15}$$

3.1.4 *Entropy model*

Shannon introduced the concept from thermodynamics called "information entropy" [81]. It is explained as the uncertainty of information calculated through the following equation [45]:

$$H(X) = -\sum_{i=1}^{n} p(a_i) \log p(a_i) \tag{16}$$

where:

X	is $a_1, a_2, \ldots, a_n$
$P(a_i)$	Represents the probability of the event a_i occurring

$$0 \leqslant p(a_i) \leqslant 1 \qquad (i = 1, 2, \ldots, n) \tag{17}$$

$$p(a_i) \qquad \text{The probability of the event } a_i \tag{18}$$

A remarkable property about entropy is the following:

$$H(p_1, p_2, \ldots, p_n) \leqslant \log n \tag{19}$$

If and only if:

$$p_i = \frac{1}{n} \qquad (i > 0) \tag{20}$$

The threshold limit for Shannon's entropy is defined as $\log(11)$ or 1.0414, bigger values show that there is sprawl in the region [73]. While values close to zero show that the growth is clustered and confined.

3.2 RESULTS

3.2.1 *Variograms*

To further improve the Kriging model, I developed variograms and later integrated ranges and sills from the variogram model, as shown in Table 3.1.

3.2.2 *Residual kriging model*

The residual Kriging model was useful in calculating the population density in 100m mesh. In the UPA of Aomori MtA (Figures 3.2, 3.3 and 3.4) it was clear that the population is concentrated in the urban core. However, there are some areas that are detached from the main part of the UPA. In the southern part there is an area with high population density. For Sendai MtA (Figures 3.5, 3.6 and 3.7), the UPA can

Table 3.1: Variogram characteristics

MtA	Model	1995		2000		2005	
		psill	range	psill	range	psill	range
Aomori	Nug	0	0	0.004	0	0.044	0
	Exp	3.898	5.688	3.85	5.844	3.722	6.672
Sendai	Nug	0.579	0	0.723	0	0.694	0
	Exp	3.024	9.729	2.768	9.424	2.637	10.009
Sapporo	Nug	0.998	0	0.992	0	0.988	0
	Exp	11.344	3,209.804	16.743	3,851.751	20.849	4,622.098

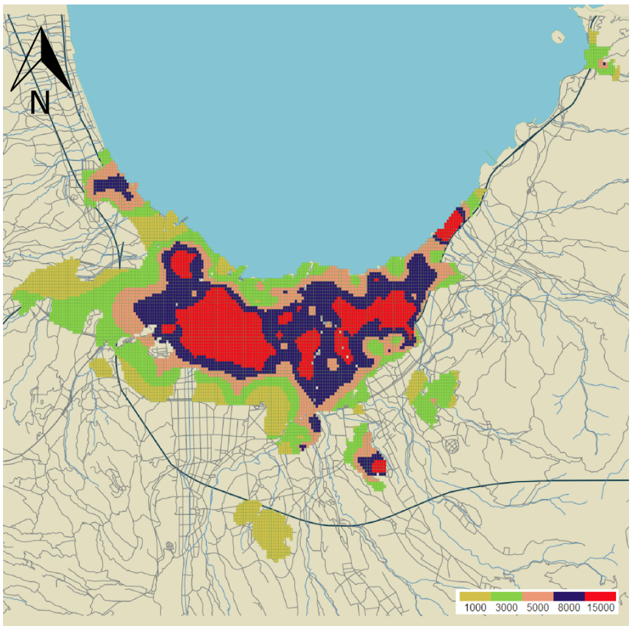


Figure 3.2: Residual kriging model for population density (people/km^2) in the UPA of Aomori MtA (1991)

be divided in two: a western area where most of residents concentrate and the commuter belt in the eastern part[54]. The commuter belt is located close to Shiogama and Shichigahama, where the Senseki line is operating and the Sen-en highway helps the transportation process to the commuter belt. For the UPA of Sapporo MtA (Figures 3.8, 3.9 and 3.10), the Kriging model allowed identification of the main area, where most of the population is concentrated.

In addition, commuter belts were identified in the areas of Tomakomai and Chitose in the southern part, and in the northern part, close to Otaru and Ishikari. By calculating the Kriging model I identified population hot spots in every period of time.

The standard deviation for the residual kriging models for every UPA in the MtA in the year 2006 is shown in the Figures 3.11, 3.12

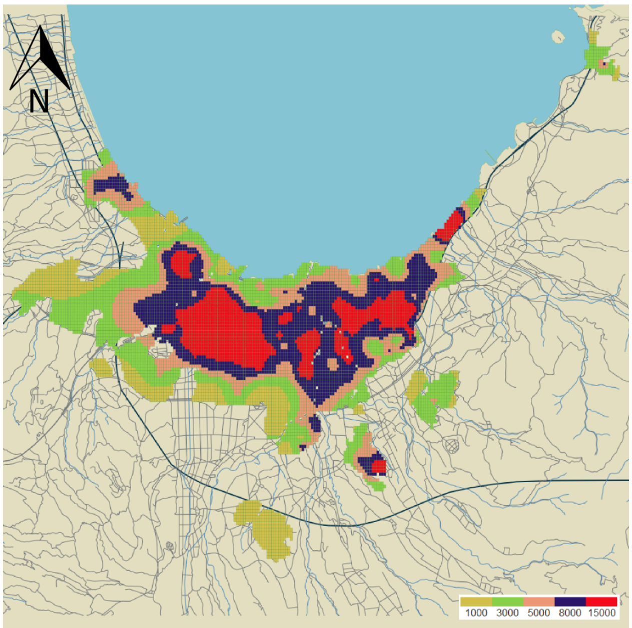


Figure 3.3: Residual kriging model for population density (people/km^2) in the UPA of Aomori MtA (1997)

and 3.13. It is shown that the minimum standard error is located in the urban core of each MtA, and outside of this the standard deviation becomes higher, for instance the UPA of Sapporo MtA in the southern part has the highest standard deviation. Because of the size of the UPA in the MtA and the scope of the model to predict population density in each 100m mesh, one can identify that the standard deviation interval for the UPA of Aomori MtA is [0.33, 089], while the UPA in Sendai and Sapporo MtAs are [0.91, 1.22] and [0.99, 2] respectively.

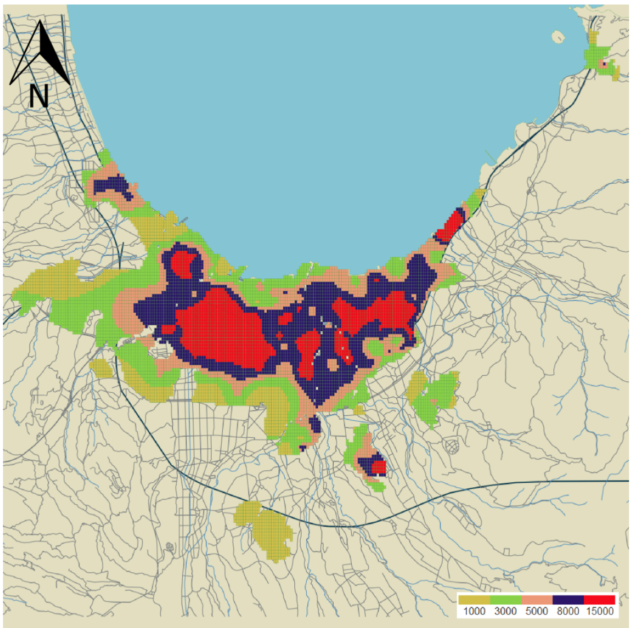


Figure 3.4: Residual kriging model for population density (people/km^2) in the UPA of Aomori MtA (2006)

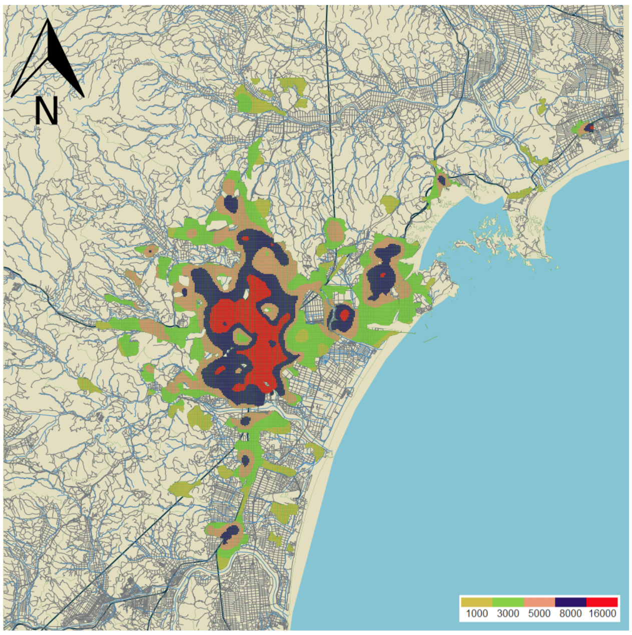


Figure 3.5: Residual kriging model for population density (people/km^2) in the UPA of Sendai MtA (1991)

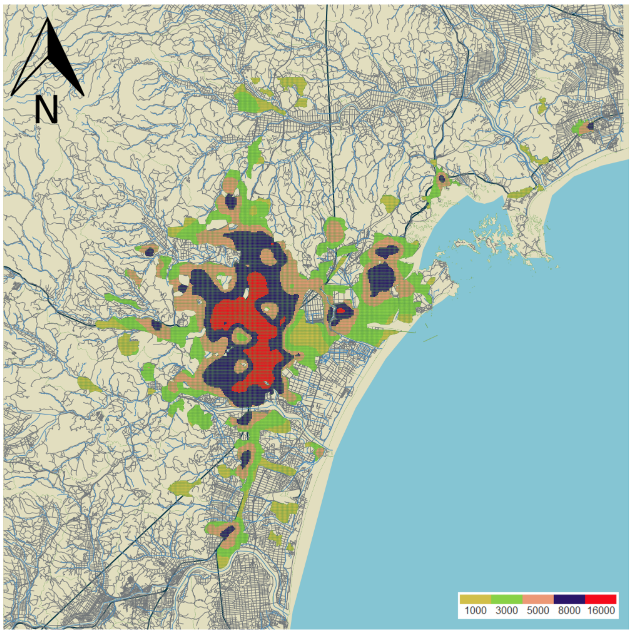


Figure 3.6: Residual kriging model for population density (people/km^2) in the UPA of Sendai MtA (1997)

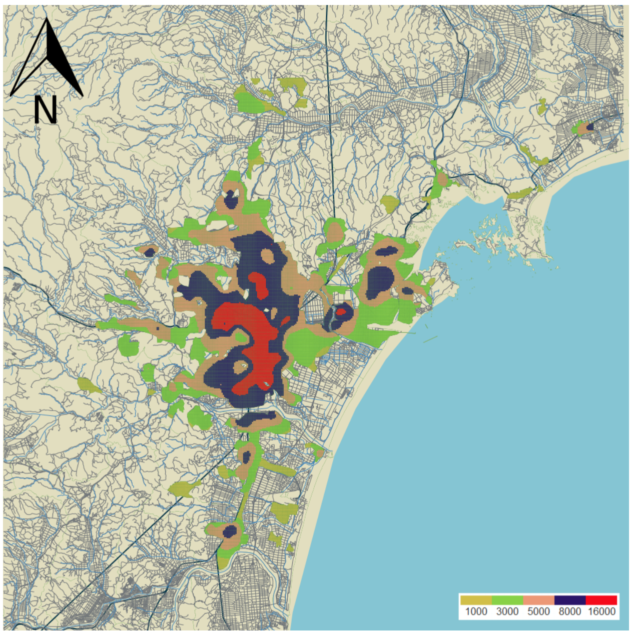


Figure 3.7: Residual kriging model for population density (people/km^2) in the UPA of Sendai MtA (2006)

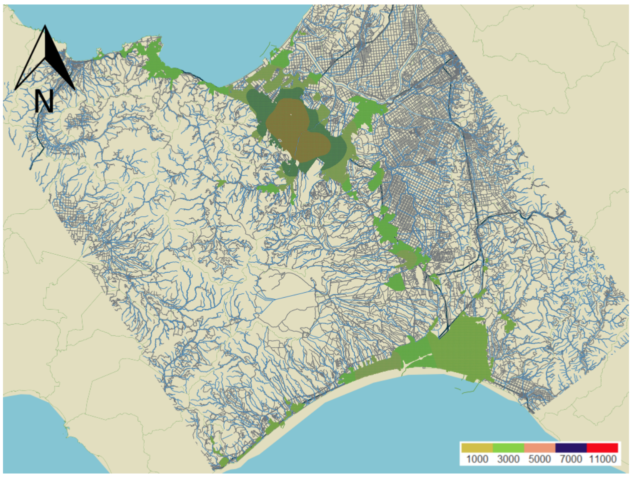


Figure 3.8: Residual kriging model for population density (people/km^2) in the UPA of Sapporo MtA (1991)

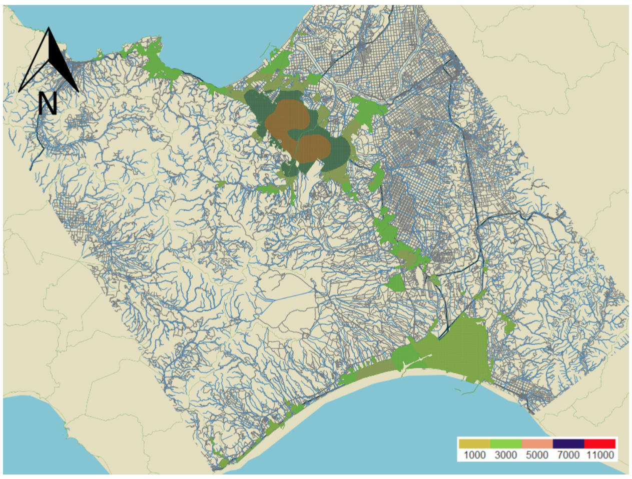


Figure 3.9: Residual kriging model for population density (people/km^2) in the UPA of Sapporo MtA (1997)

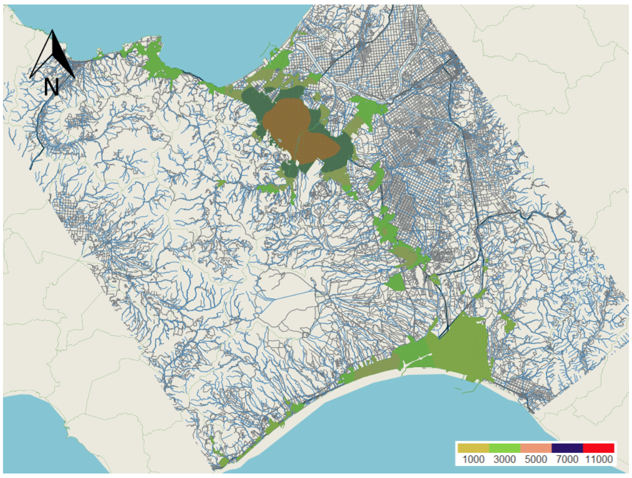


Figure 3.10: Residual kriging model for population density (people/km^2) in the UPA of Sapporo MtA (2006)

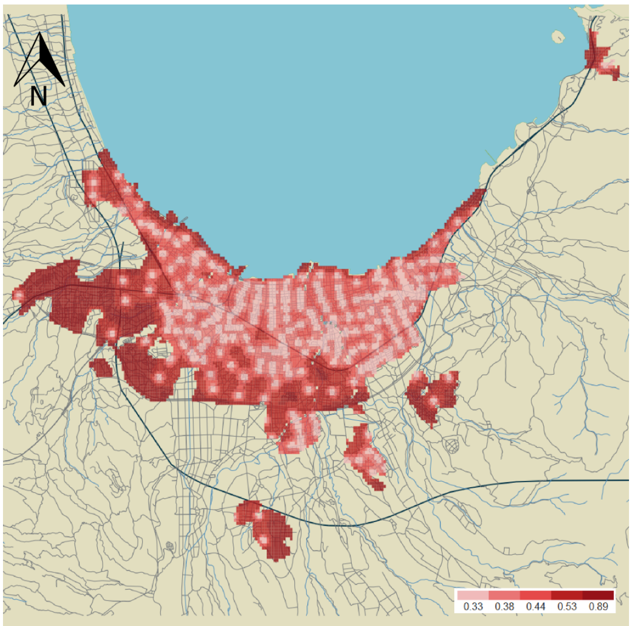


Figure 3.11: Standard deviation for the UPA of Aomori MtA (2006)

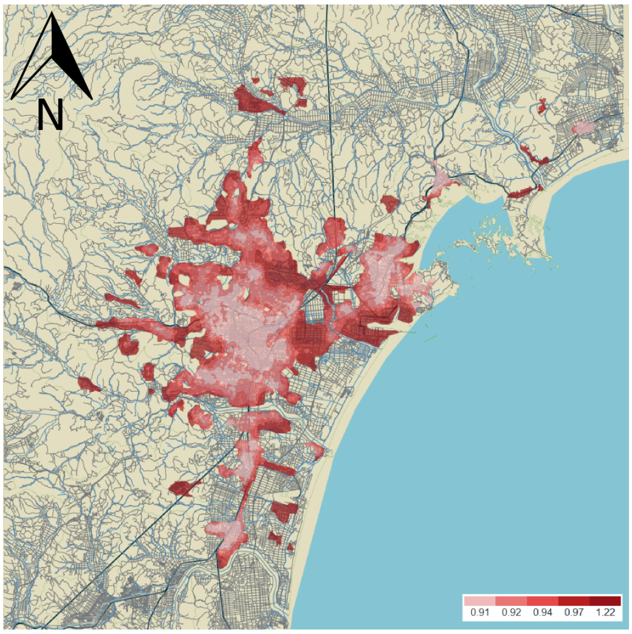


Figure 3.12: Standard deviation for the UPA of Sendai MtA (2006)

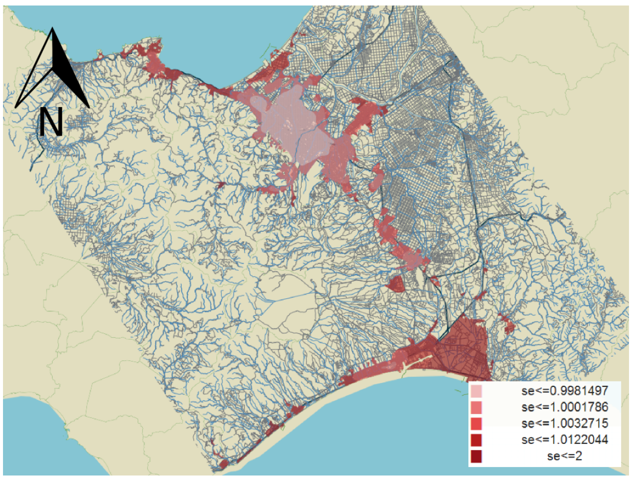


Figure 3.13: Standard deviation for the UPA of Sapporo MtA (2006)

Table 3.2: Compactness calculation for each UPA in their MtAs

city	year	densblt	densres	supfacs	dens	mixuse	compact	code
Aomori	1991	1.27	1.57	0.13	1.42	0.13	0.78	a91
Aomori	1997	1.35	1.36	0.58	1.36	0.58	0.97	a97
Aomori	2006	1.31	0.79	1.80	1.05	1.80	1.42	a06
Sendai	1991	-0.53	0.06	-1.07	-0.24	-1.07	-0.65	sd91
Sendai	1997	-0.42	-0.48	-0.26	-0.45	-0.26	-0.36	sd97
Sendai	2006	-0.37	-0.83	0.97	-0.60	0.97	0.19	sd06
Sapporo	1991	-0.97	-0.57	-1.29	-0.77	-1.29	-1.03	sp91
Sapporo	1997	-0.82	-0.80	-0.76	-0.81	-0.76	-0.79	sp97
Sapporo	2006	-0.81	-1.09	-0.11	-0.95	-0.11	-0.53	sp06

3.2.3 *Compactness*

The compactness index (compact) was calculated and is shown in Table 3.2. As expected, the most compact area corresponds to the UPA of Aomori MtA in the year 2006, with a higher index (1.42) than other MtAs. The UPA of Aomori MtA is the smallest one in this study and its compactness index values for all the different periods of time are positive. The UPA of Sendai MtA is a mid-sized metropolis, and it presents a positive compactness index only in the latest period of time. This demonstrates that although residential areas in the UPA are being promoted, it is necessary to improve this in order to increment the index of compactness in this MtA.

The UPA of Sapporo MtA also presents an increment, which is not as fast as the other MtAs. This area is the largest MtA among the study areas. In the same way as in other MtAs, it is necessary to improve residential areas. Although residential areas increase in the core of the UPA, they also increase in the commuter belt. If the situation continues in this direction, it will be difficult to reach an appropriate compact city model in this MtA.

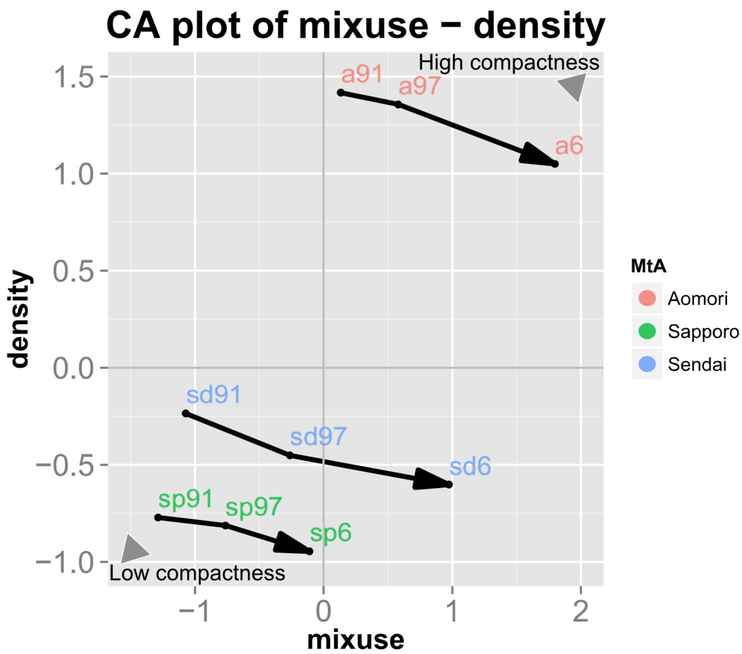


Figure 3.14: Correspondence analysis between Mixuse and Density for all MtAs

Table 3.3: Entropy values for each MtA

	UPA	Year	Entropy	Code
1	Aomori	1991	1.609	a091
2	Aomori	1997	1.512	a097
3	Aomori	2006	1.286	a006
4	Sendai	1991	1.767	sd91
5	Sendai	1997	1.618	sd97
6	Sendai	2006	1.411	sd06
7	Sapporo	1991	1.839	sp91
8	Sapporo	1997	1.744	sp97
9	Sapporo	2006	1.559	sp06

3.2.4 *Entropy*

The entropy was calculated for every MtA as shown in Table 3.3. The MtA presenting less entropy for the land use was the UPA of Aomori MtA in the period 2006, however this value is still above the entropy's threshold. It can be seen the efforts of each local government to promote the UPA as a residential area. Figure 3.15 shows a similar behavior in each MtA.

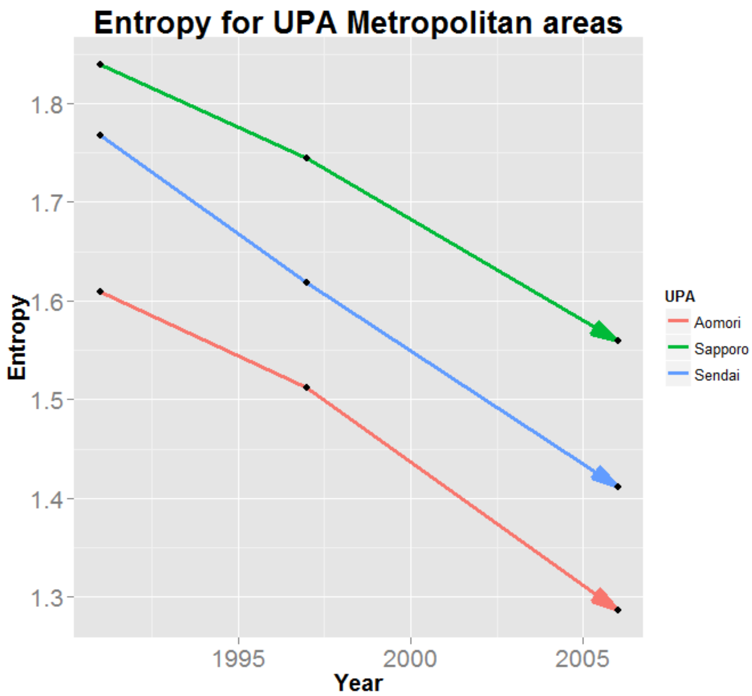


Figure 3.15: Entropy behavior for MtAs

Figures (3.16, 3.17 and 3.18) show a simulation for each UPA in their MtAs. The simulation was performed by calculating the multinomial distribution for each MtA in each period of time. The probability function is represented as equation 21. 10,000 simulations for each MtA in every period were calculated, thus the behavior of the MtA was known. Since the UPA is specially designated for residential area this subsection is focused on this type of land use. In the UPA of Aomori MtA in the year 1991, it is evident the difficulty for the residential area to reduce the entropy in 0.1 points, also the entropy values between the periods 1991 and 1997 were closer. In 2006 the entropy was reduced by more than 0.3 points since 1991. The behavior of the UPA of Sapporo and Sendai MtAs is similar, although the entropy's value for the UPA of Sapporo MtA remains higher. The entropy in the last period of time is at least 2.5 points less than 1991's value. It suggests that the local governments promoted the residential area rather than another type of land use.

$$P(X_1 = x_1, \cdots, X_n = x_n) = \frac{N!}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \theta_i^{x_i} \tag{21}$$

Where x_i are nonnegative integers, and θ_i are constants with $\theta_i > 0$.

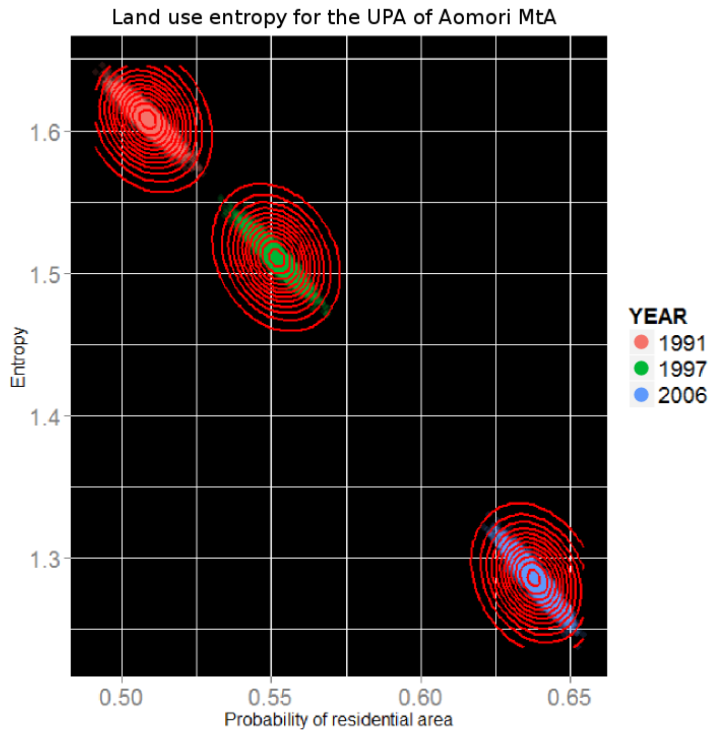


Figure 3.16: Simulation of entropy for land use in the UPA of Aomori MtA

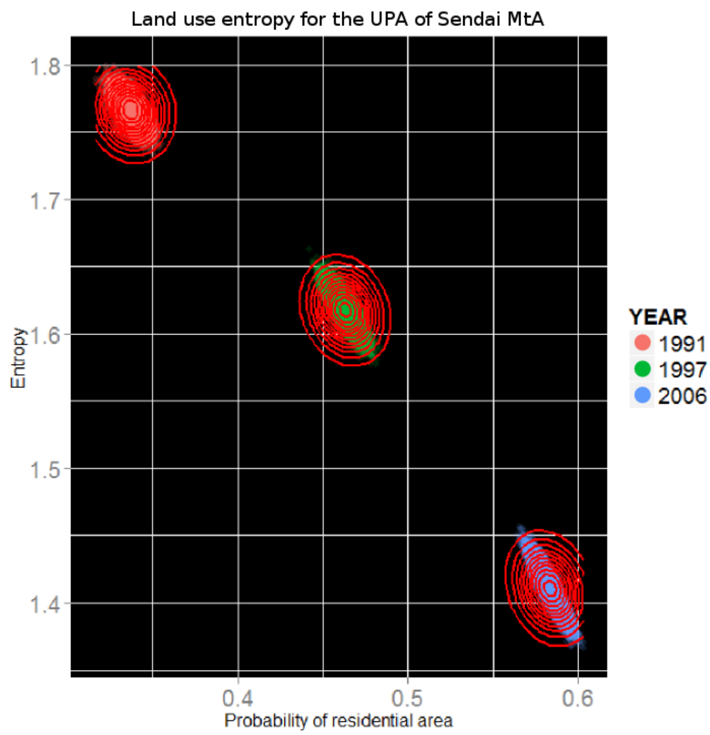


Figure 3.17: Simulation of entropy for land use in the UPA of Sendai MtA

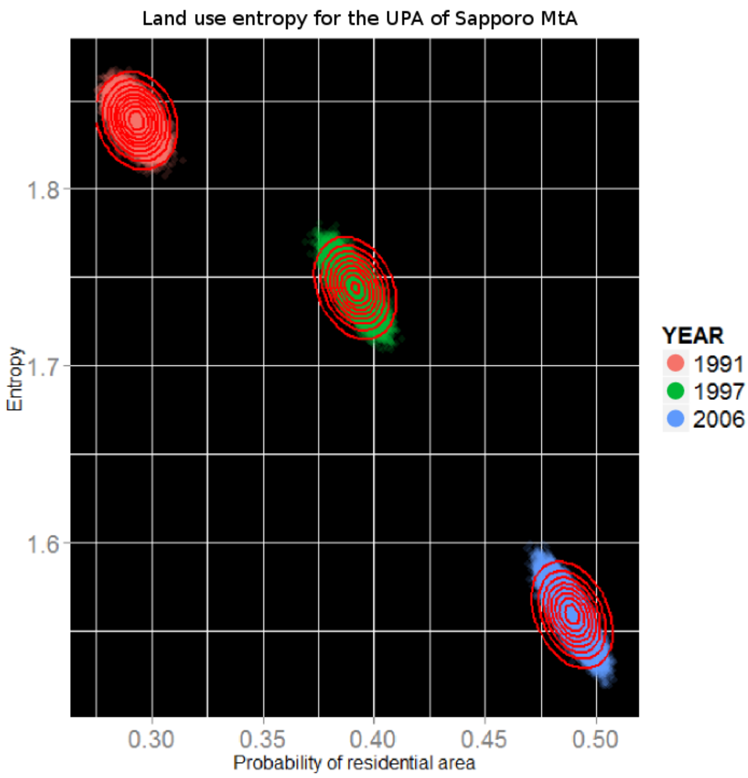


Figure 3.18: Simulation of entropy for land use in the UPA of Sapporo MtA

4 EVALUATION FOCUSING ON SOCIO-ECONOMIC FACTORS AND LAND USE

In this chapter, I will focus on the second basic condition of the compact city model, this is the one related to socio-economic factors. One of the goals of the compact city model is the activity concentration. This is linked to the daily facilities such as the ones prepared to provide social (banks, schools, universities, convenience stores, etc.), health (hospitals, private clinics), meals (restaurants) and therapeutic services (private and public clinics, etc.), and they should be well distributed in the urban area in order to reduce traveling time. Also, the public transportation intensification such as bus stops is required to reduce the CO_2 emissions and thus avoid traffic congestions.

The housing decision process should be taken carefully because it will affect the behavior and daily routine. This process is different for each person, however a collective knowledge could be extracted studying the population settlement. Then, a question appears: How to find a particular place maximizing user utilities?. In this chapter a supervised non-parametric statistical technique is applied to understand how the land use types can be classified using socio-economic factors; in this way the SVM applied for classification will be introduced.

First of all, I will explain the methodology with a special focus on parallel computing, because it was a necessary step forward for classifying a large UPA in a MtA .

4.1 METHODOLOGY

4.1.1 *Haversine distance calculation*

Using GIS, I calculated the haversine distance for each polygon to the different socio-economic factors, the area of each polygon is 1 ha. The haversine distance is defined as follows:

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\left(\varphi_1\right)^* \cos\left(\varphi_2\right) \sin^2\left(\frac{\Delta\lambda}{2}\right) \tag{22}$$

$$c = 2\tan 2(\sqrt{a}, \sqrt{1-a}) \tag{23}$$

$d = R * c$
where:

φ Latitude
λ Longitude
R The earth's radius, defined here as (6,371km)
d The haversine distance

4.1.2 *Land prices*

The price of land is measured by district and it was necessary to calculate it by 100m mesh areas. It can be defined as:

$$Pr_a = \sum_{j=1}^{n} A_{a \in j} Pr_j \tag{24}$$

Where:
Pr_a The price defined for the polygon a
j The districts intercepted by polygon a
$A_{a \in j}$ The area that shares a with the district j
Pr_j The price defined by the district j

However, the land use price is taken as the price defined by the district where the mesh area lies.

4.1.3 *Support Vector Machine*

The SVM is a supervised non-parametric statistical learning technique, which has an important property: The determination of the model parameters corresponds to a convex optimization problem, where a local solution is also a global optimum.

Recently, SVM is applied in different kind of studies such as analyze data, recognize patterns, classification and regression analysis. For instance Zhou et al. [103] presented a method of Japanese dependency structure analysis based on SVM and conditional random fields (CRF). Their experiments demonstrated that combining SVM and CRF outperforms the cascaded chunking model based on sole SVMs and sole CRFs. Sudha and Bhavani [90] compared the efficiency between the k-Nearest Neighbor models (kNN) and multi class SVMs where multiple gait components were fused for enhancing the classification rate. Their results demonstrated that the classification method using SVM was better than kNN. Pitiranggon et al. [69] developed a decompositional rule extraction technique from SVM called Support Vector Space Expansion (SVSE) rule. They applied it to financial data to predict currency crises. Ramirez-Gutierrez et al. [74] developed a face recognition algorithm using eigenphases and histogram equalization. They proposed a featured extraction scheme using SVM, the recognition rate was higher than 97% and verification error lower than 0.003%.

Given a two separable classes with k samples defined as (x_i, y_i), where $i = 1, 2, \ldots, k$, where $x \in R^n$ is an n-dimensional space, and $y_i \in \{+1, -1\}$ is a class label [38]. Suppose that the classes could be separated by two hyperplanes parallel to the optimal hyperplane.

$$w \cdot x_i + b \geqslant 1 \qquad \text{for } y = 1 \text{ and } i = 1, \ldots, k. \tag{25}$$

$$w \cdot x_i + b \leqslant 1 \qquad \text{for } y = -1 \tag{26}$$

Where $w = (w_1, \ldots, w_n)$ is a vector of n elements.

Equations (25) and (26) can be combined into a single equation:

$$y_i \left[w' x_i + b\right] \geqslant 1, \text{ where } i = 1, \ldots, k. \tag{27}$$

The training data points on the hyperplanes that are parallel to the Optimal Separated Hyperplane (OSH) are the support vectors. The margin between the planes is defined as: $\frac{2}{|w|}$, the generalized linear SVM find an OSH by solving the Figure 4.1. The optimization problem can be written as follows:

$$\min \left[\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{r} \xi_i\right] \tag{28}$$

Subject to:

$$y_i \left[w \cdot x_i + b\right] + \xi_i - 1 \geqslant 0 \tag{29}$$

$$\xi_i \geqslant 0 \tag{30}$$

Where C is a penalty parameter on the training error, and ξ_i is the non-negative slack variable. The optimization model can be solved by introducing Lagrange multipliers for its dual optimization model [64]. If it is not possible to find a hyperplane by linear equations, the data must be mapped into a high dimensional space using nonlinear mapping functions (Φ).

$$f(x) = sgn\left(\sum_{i=1}^{k} \alpha_i y_i k(x, x_i) + b\right) \tag{31}$$

Where α_i is a Lagrange multiplier. $k(x, x_i)$ is a positive kernel that must meet Mercer's condition. Kernel functions can be aggregated into linear, polynomial, radial basis (Gaussian) functions and sigmoid kernels.

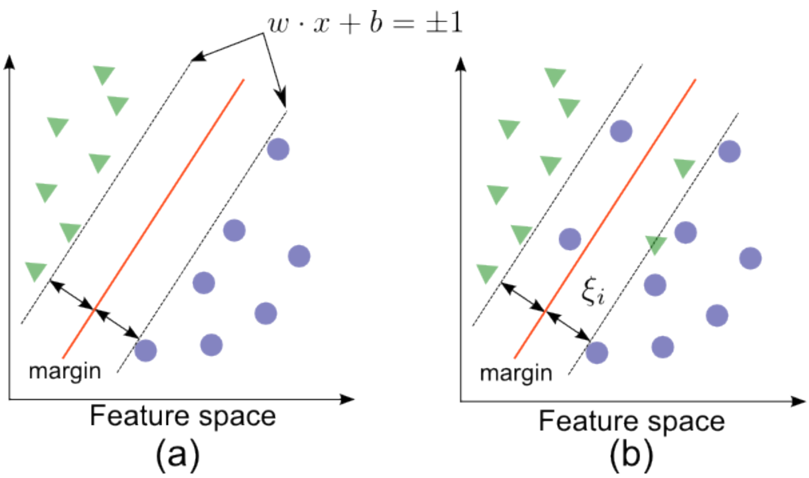


Figure 4.1: Optimal separating hyperplane. (a) separable samples, (b) non-separable data samples.

$$\text{Linear kernel}: k(x_i, x) = (x_i, x) \tag{32}$$

$$\text{Polynomial kernel}: k(x_i, x) = (x_i \cdot x + 1)^d \tag{33}$$

where d is a natural number

$$\text{Radial basis function kernel}: k(x_i, x) = exp\left(-\frac{1}{\sigma^2}\|x_i - x\|^2\right) \tag{34}$$

$$\text{Sigmoid kernel}: (x_i, x) = \tanh(kx_i \cdot x - \delta) \tag{35}$$

By using the kernel function the nonlinear SVM classifier is defined as:

$$\text{sign}\left(\sum_{i=1}^{k} \alpha_i^* y_i k(x_i, x) + b^*\right) \tag{36}$$

SVMs were developed for binary classifications, however several studies such as one-against-all, one-against one and all together have been done for multiple class classification scenarios. In one-against-all a set of binary classifiers, each trained to separate one class from the rest [58]. One-against-one approach $k(k-1)/2$ SVMs are constructed for each pair of classes, the training data vector x_i is predicted to belong to the class with maximum number of votes.

The experiments of Yu and Agouris [102] give an approach of spatial information management, so I have defined the support vector machine in the following way:

$$F(\vec{lu})_i = (\alpha * lat_i + \beta * long_i) + \sum_{j=1}^{m} \gamma_j * dist_{ij} + \zeta * price_i \tag{37}$$

Where:

i	Mesh area i
m	Total number of socio-economic factors classes
lat	Latitude
long	Longitude
$dist_{ij}$	Minimum distance from mesh area i to socio-economic factor type j
$price_i$	Price of land use in which mesh i lies
$\alpha, \beta, \gamma, \zeta$	Coefficients
$F(\vec{lu})$	Classification function between all land use types

4.1.4 *Area under the curve (AUC)*

The receiving operating characteristics (ROC) allows assessing uncalibrated decision functions, even when the prior distribution of classes is unknown. The ROC curve details the rate of true positives against false positives over a threshold. The Area of the ROC curve is the probability that a randomly drawn positive example has a higher decision function value than a negative example, it is called the AUC. The AUC is close to the Gini coefficient, the last one is used in random forest algorithms for classification variables.

The AUC is defined as follows:

$$AUC_{Total} = \frac{2}{\|C\|(\|C\|-1)} \sum_{c_i, c_j \in C}^{n} AUC(c_i, c_j) \tag{38}$$

Where n is the number of classes, $AUC(c_i, c_j)$ is the area under the two-class ROC curve between the classes c_i and c_j, and C is the set of all classes. The execution time of SVM [37] is calculated as:

$$T_{ex} = O\left(|C|^2 n \log n\right) \tag{39}$$

Where:

n	Number of elements
C	Number of classes

4.1.5 *Cross-validation method*

The cross-validation method is a model evaluation method that is better than simply looking at the residuals [40]. This method tends to focus on a sample data, that data is called training set and the remaining data is called testing set. Once the model has been elaborated it is possible to compare the predicted data using the training set with the testing set through equation (41).

$$\tilde{y} = A \cdot (y_1, \ldots, y_{i-1}, \tilde{y}_i, \tilde{y}_{i+1}, \tilde{y}_N)^T \tag{40}$$

$$CV(\widehat{f}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \tilde{y}_{\delta i})^2 \tag{41}$$

Where:

y_i	Input data
$\tilde{y}$	A combination of y_i
$y_i - \tilde{y}_{\delta i}$	Denominated as noise

4.2 RESULTS

4.2.1 *UPA of Aomori MtA*

By calculating and rescaling the distance from each polygon's centroid to the socio-economic factors, it was possible to evaluate the geographical distribution. Figure 4.2 shows a density curve for distance from each polygon to socio-economic factors. According to the compact city definition, the boundaries have to be defined clearly; although the edges in the main part of the UPA are not defined particularly, it is possible to identify using geographical data that most of the social activities take place in this area. By calculating distances, it was possible to identify the boundaries of the UPA central part at 5km from the core, further than this distance there are few railway stations, supermarkets and other facilities.

4.2.1.1 *Parameterization*

First of all, the database was randomly divided for each MtA into two sets, 70% of the data was used as training and the remaining information (30%) was used for testing. The training set is used to

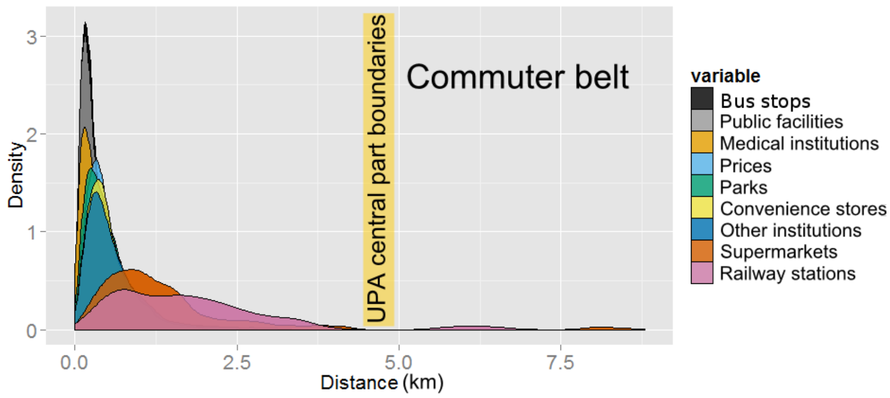


Figure 4.2: Distance from polygon to socio-economic factors (km)

train a multiclass SVM classifier [3]. According to literature on land use prediction models using SVM [84][46], the radial basis function kernel presents a good performance in land use prediction. For that reason the SVM has been configured using this kernel distribution.

I have performed 3 different experiments to find a minimum training and cross errors, those experiments are called tuning 1, tuning 2 and best model. The parameters related to cost and gamma values are determined by a grid search method using cross validation approach. The grid search method is useful for the computation of expensive numerical simulations and it has been applied in different studies in order to find a global minimum [61].

The first experiment consisted of 30 simulations, and total time was 19.90 min. The Kappa and Rand indexes were calculated to evaluate the SVM, the results were (0.47, 0.71) respectively. The Kappa index with larger values indicates better reliability, Kappa values greater than 0.7 are considered satisfactory[1]. The Rand index measures the percentage of decisions that are correct[2]; it means that the prior results are still far from an accurate value. For that reason the experiments have to be improved. The number of vectors for this experiment was 2,386.

The computational time for the second experiment was 93.30 min. In this experiment, I have run 200 simulations; the Kappa and Rand indexes were (0.63, 0.79) respectively, and the percentage of diagonal values was 81%, however the results are still far from good accuracy levels. The total number of vectors was 2,537. For those reasons I have extended the grid search in order to reach more accuracy and best model parameters.

The final experiment took more than 39 hours using a Windows based computer with 12Gb RAM memory, and processor Intel Xeon 2.67 Ghz. The total number of simulations was 5,015. In this experiment I found appropriate Kappa and Rand Indexes (0.82, 0.86) respectively, the percentage of diagonal values is 90%. The results of the experiments are shown in Table 4.1. In this experiment, the cost value same as 100 presents the minimum training error (10.0%) when the cross value is same as 30 (Figure 4.3), Table 4.2 presents simulation sample values. The number of vectors calculated was 2,429 with the same number of predictors. Although the first experiment presents the minimum number of vectors, the third experiment shows the best Kappa and Rand indexes.

The largest errors were on waste land (21.4%), across rivers and lakes (25.6%). However, the error close to beach area was (0.0%), forest (6.4%), buildings and residential area (6.7%).

---

1 Cohen's Kappa: http://psych.unl.edu/psycrs/handcomp/hckappa.PDF, last access: August 22, 2014.

2 Evaluation of clustering: http://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html#24376, last access: August 22, 2014.

Table 4.1: Experimental results

	Tuning 1	Tuning 2	Best model
Gamma value	0.01	0.1	0.07
Cost	760	7	100
Kappa Index	0.47	0.63	0.82
Rand Index	0.71	0.79	0.86
% diagonal values	0.74	0.81	0.90
Processing time (min)	19.9	93.75	>12 (hours)

Table 4.2: Aomori SVM results

ID	**Cost**	**Cross**	**Sigma**	**NSV**	**Trer**	**Crer**
93	100	30	0.0684	2,429	0.1003	0.2587
96	100	60	0.0682	2,429	0.1003	0.2624
90	90	100	0.0717	2,431	0.1008	0.2625
91	100	10	0.0679	2,428	0.1010	0.2653
94	100	40	0.0676	2,427	0.1012	0.2660
95	100	50	0.0674	2,429	0.1015	0.2639
97	100	70	0.0674	2,427	0.1019	0.2627
98	100	80	0.0674	2,428	0.1019	0.2655
99	100	90	0.0665	2,426	0.1022	0.2648
87	90	70	0.0693	2,436	0.1029	0.2615
84	90	40	0.0691	2,430	0.1033	0.2644
89	90	90	0.0689	2,429	0.1033	0.2635
86	90	60	0.0688	2,429	0.1036	0.2630
81	90	10	0.0679	2,427	0.1043	0.2688
83	90	30	0.0679	2,427	0.1043	0.2649
88	90	80	0.0675	2,426	0.1048	0.2660
85	90	50	0.0673	2,425	0.1052	0.2608
79	80	90	0.0701	2,427	0.1073	0.2618
75	80	50	0.0688	2,420	0.1083	0.2630
82	90	20	0.0655	2,419	0.1085	0.2670
78	80	80	0.0686	2,419	0.1090	0.2627
73	80	30	0.0684	2,422	0.1095	0.2660
72	80	20	0.0682	2,424	0.1097	0.2608
74	80	40	0.0683	2,423	0.1097	0.2615
80	80	100	0.0679	2,421	0.1102	0.2628
76	80	60	0.0678	2,422	0.1106	0.2632
71	80	10	0.0668	2,419	0.1125	0.2688
65	70	50	0.0703	2,425	0.1132	0.2648
77	80	70	0.0657	2,416	0.1144	0.2608

Continued on next page

Table 4.2 – *Continued from previous page*

ID	Cost	Cross	Sigma	NSV	Trer	Crer
68	70	80	0.0690	2,422	0.1158	0.2616
70	70	100	0.0688	2,422	0.1161	0.2626
61	70	10	0.0684	2,426	0.1168	0.2646
67	70	70	0.0683	2,428	0.1170	0.2611
63	70	30	0.0679	2,429	0.1177	0.2604
69	70	90	0.0680	2,429	0.1177	0.2627
54	60	40	0.0705	2,430	0.1208	0.2601
62	70	20	0.0654	2,440	0.1222	0.2587
64	70	40	0.0651	2,439	0.1229	0.2629
66	70	60	0.0652	2,438	0.1229	0.2625
52	60	20	0.0691	2,438	0.1243	0.2674
53	60	30	0.0684	2,440	0.1255	0.2575
60	60	100	0.0682	2,442	0.1262	0.2640
56	60	60	0.0679	2,442	0.1266	0.2599
55	60	50	0.0677	2,441	0.1271	0.2611
57	60	70	0.0676	2,442	0.1271	0.2608
51	60	10	0.0670	2,436	0.1278	0.2672
59	60	90	0.0667	2,436	0.1278	0.2622
58	60	80	0.0658	2,431	0.1299	0.2639
46	50	60	0.0703	2,442	0.1318	0.2644
41	50	10	0.0685	2,438	0.1344	0.2599
44	50	40	0.0685	2,439	0.1344	0.2615
45	50	50	0.0675	2,442	0.1349	0.2594
49	50	90	0.0681	2,440	0.1349	0.2622
42	50	20	0.0679	2,442	0.1351	0.2608
43	50	30	0.0680	2,440	0.1351	0.2618
48	50	80	0.0672	2,441	0.1356	0.2622
47	50	70	0.0667	2,444	0.1368	0.2592
50	50	100	0.0660	2,452	0.1372	0.2608
39	40	90	0.0729	2,451	0.1379	0.2599
36	40	60	0.0702	2,464	0.1427	0.2620
33	40	30	0.0682	2,468	0.1434	0.2622
38	40	80	0.0683	2,468	0.1434	0.2649
35	40	50	0.0678	2,475	0.1436	0.2646
37	40	70	0.0680	2,470	0.1436	0.2601
31	40	10	0.0688	2,466	0.1438	0.2601
32	40	20	0.0673	2,477	0.1445	0.2606
34	40	40	0.0663	2,476	0.1460	0.2616
40	40	100	0.0650	2,474	0.1488	0.2627
25	30	50	0.0703	2,473	0.1514	0.2665
23	30	30	0.0691	2,475	0.1535	0.2660
29	30	90	0.0690	2,475	0.1542	0.2660
22	30	20	0.0685	2,476	0.1549	0.2683

Continued on next page

Table 4.2 – *Continued from previous page*

ID	Cost	Cross	Sigma	NSV	Trer	Crer
27	30	70	0.0669	2,480	0.1577	0.2628
21	30	10	0.0666	2,479	0.1589	0.2622
28	30	80	0.0662	2,477	0.1596	0.2656
26	30	60	0.0654	2,475	0.1620	0.2609
30	30	100	0.0646	2,474	0.1631	0.2643
24	30	40	0.0633	2,473	0.1653	0.2629
13	20	30	0.0706	2,473	0.1740	0.2611
17	20	70	0.0695	2,478	0.1744	0.2586
12	20	20	0.0687	2,478	0.1747	0.2636
14	20	40	0.0685	2,477	0.1747	0.2599
16	20	60	0.0681	2,479	0.1751	0.2611
18	20	80	0.0678	2,478	0.1754	0.2604
19	20	90	0.0679	2,478	0.1754	0.2585
11	20	10	0.0670	2,484	0.1777	0.2672
15	20	50	0.0668	2,486	0.1777	0.2611
20	20	100	0.0670	2,485	0.1777	0.2598
2	10	20	0.0699	2,495	0.2020	0.2667
8	10	80	0.0691	2,499	0.2029	0.2653
1	10	10	0.0683	2,501	0.2041	0.2618
7	10	70	0.0678	2,501	0.2046	0.2656
3	10	30	0.0677	2,503	0.2048	0.2679
5	10	50	0.0674	2,502	0.2050	0.2660
6	10	60	0.0673	2,501	0.2053	0.2647
10	10	100	0.0668	2,500	0.2060	0.2663
9	10	90	0.0664	2,501	0.2067	0.2652
4	10	40	0.0663	2,502	0.2072	0.2642

Where:

ID Exercise ID
Cost Cost value
Cross Cross value
Sigma Sigma value
NSV Number of supported vectors
Trer Training error
Crer Cross error

In this chapter the same nomenclature will be used for all the tables related with SVM results.

4.2.1.2 *AUC calculation*

The third experiment was used to calculate the AUC for each predictor (Figure 4.4), the highlighted title represents the land use types.

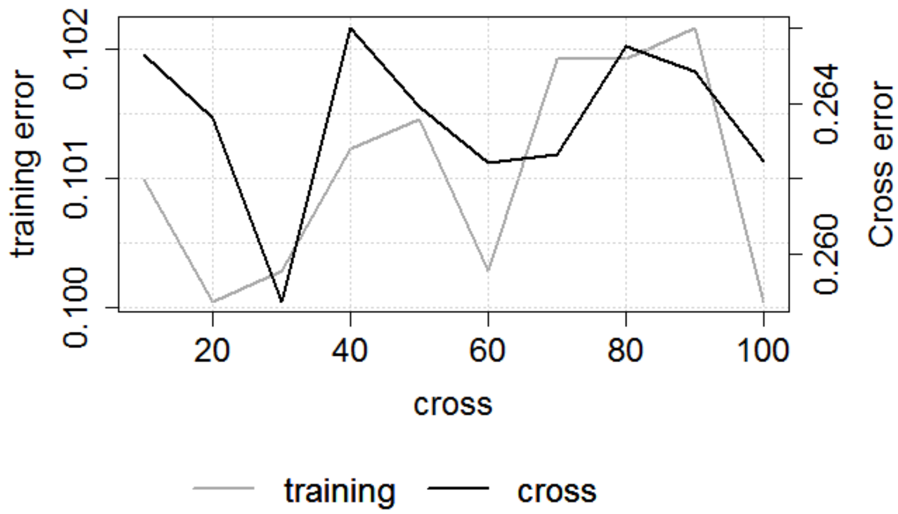


Figure 4.3: Results with cost same as 100 for the UPA of Aomori MtA

The results for building and residential area parameter show that latitude is the factor that contributes most to the model with an area higher than (0.95). The factors given by the distance to other institutions, public facilities, medical institutions, longitude and convenience stores conformed the second group with an AUC value higher than (0.8). Finally, in a third group there are factors related with transportation methods, such as distance to train stations and bus stops, also the public facilities categories, parks, supermarkets. For all the factors the "latitude" has the largest AUC with a value higher than 0.9. And price of land is the factor that aggregates less information to the model.

Distance to the closest area where the prices of land use have been defined is below 0.8 as well as distance to supermarkets, and prices, the socio-economic factor with lower AUC for all the types of land use corresponds to the type of public facility.

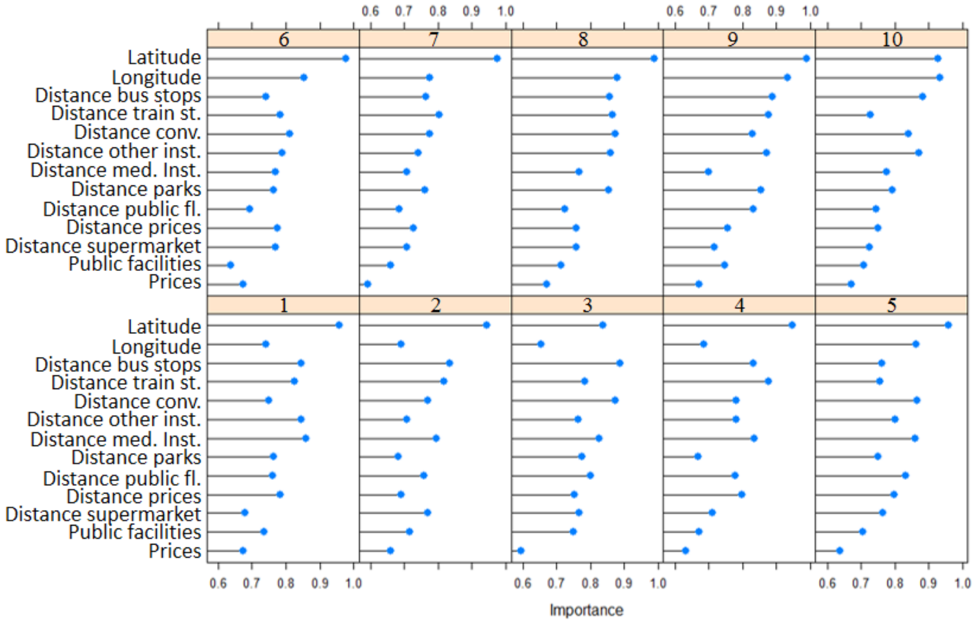


Figure 4.4: AUC for each predictor

Table 4.3: Land use categories

Old code	Description	New code	Color
1	Rice fields	a	
2	Other agricultural land	a	
3	Forest	c	
4	Waste Land	d	
5	Buildings, residential areas	e	
6	Roads	f	
7	Other sites	A	
8	Rivers and lake areas	B	
9	Beach	E	
10	Ocean	F	
11	Golf course	G	

4.2.2 *UPA of Sendai MtA*

According to the SVM experiments in the UPA of Aomori MtA, it is possible to reduce the number of land use factors. In Table 4.3 the column "new code" shows the renamed factors. Figure 4.5 shows the process for SVM calculation, it can be described as follows:

1. Loading socio-economic factors database.

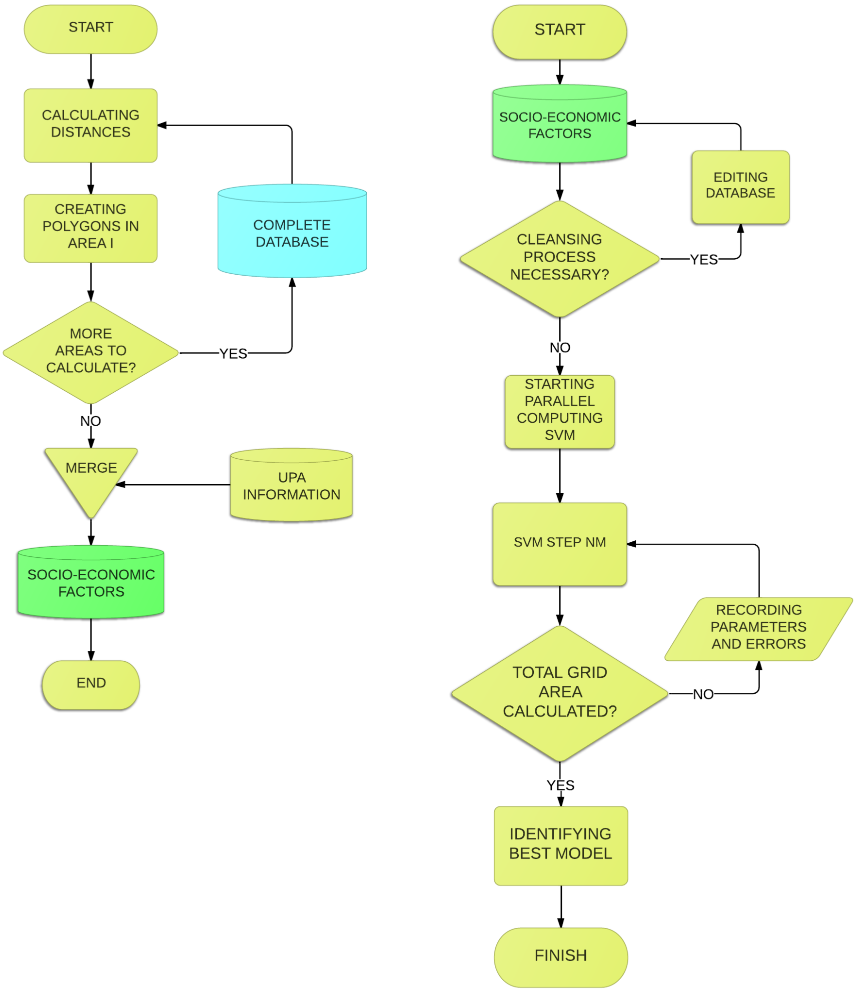


Figure 4.5: SVM procedure calculation for the UPA of Sendai MtA

2. Apply a cleansing process to the database in order to avoid future errors.
3. If the process of cleansing is not necessary the SVM calculation starts.
4. Scan the grid area using the SVM and record parameters and errors.
5. The identification of the best model is done using training and cost errors.

The results of the experiments in the UPA of Sendai MtA are shown in Table 4.4. They show that when the cost was 100 the training error was minimum (Figure 4.6).

Table 4.4: Sendai SVM results

ID	Cost	Cross	Sigma	NSV	Trer	Crer
92	100	20	0.1110	12,209	0.1828	0.2556
91	100	10	0.1102	12,212	0.1833	0.2570
97	100	70	0.1101	12,212	0.1833	0.2556
100	100	100	0.1102	12,211	0.1833	0.2556
96	100	60	0.1099	12,213	0.1835	0.2544
98	100	80	0.1092	12,222	0.1839	0.2547
95	100	50	0.1084	12,215	0.1849	0.2571
99	100	90	0.1080	12,215	0.1854	0.2557
86	90	60	0.1111	12,233	0.1857	0.2548
93	100	30	0.1071	12,211	0.1861	0.2558
85	90	50	0.1105	12,236	0.1862	0.2559
82	90	20	0.1103	12,237	0.1862	0.2570
90	90	100	0.1104	12,235	0.1862	0.2548
81	90	10	0.1099	12,240	0.1863	0.2589
84	90	40	0.1096	12,238	0.1864	0.2573
94	100	40	0.1061	12,213	0.1864	0.2577
83	90	30	0.1094	12,236	0.1866	0.2546
89	90	90	0.1089	12,236	0.1873	0.2565
80	80	100	0.1131	1,2264	0.1877	0.2568
88	90	80	0.1085	12,239	0.1879	0.2550
87	90	70	0.1082	12,240	0.1883	0.2558
76	80	60	0.1120	12,250	0.1891	0.2558
78	80	80	0.1117	12,248	0.1892	0.2563
79	80	90	0.1110	12,238	0.1899	0.2555
72	80	20	0.1107	12,243	0.1900	0.2584
75	80	50	0.1095	12,256	0.1907	0.2568
77	80	70	0.1097	12,262	0.1908	0.2557
73	80	30	0.1090	12,252	0.1911	0.2578
71	80	10	0.1088	12,252	0.1914	0.2628
74	80	40	0.1082	12,261	0.1919	0.2570
61	70	10	0.1118	12,278	0.1931	0.2584
67	70	70	0.1111	12,268	0.1940	0.2574
69	70	90	0.1107	12,265	0.1945	0.2579
65	70	50	0.1106	12,265	0.1946	0.2587
62	70	20	0.1102	12,268	0.1950	0.2593
64	70	40	0.1093	12,271	0.1957	0.2591
68	70	80	0.1093	12,271	0.1957	0.2571
66	70	60	0.1091	12,271	0.1959	0.2579
63	70	30	0.1091	12,275	0.1960	0.2580
70	70	100	0.1091	12,272	0.1960	0.2586
52	60	20	0.1120	12,315	0.1980	0.2607

Continued on next page

Table 4.4 – *Continued from previous page*

ID	Cost	Cross	Sigma	NSV	Trer	Crer
59	60	90	0.1111	12,310	0.1986	0.2587
58	60	80	0.1107	12,317	0.1990	0.2590
56	60	60	0.1106	12,314	0.1990	0.2592
60	60	100	0.1106	12,321	0.1991	0.2589
54	60	40	0.1104	12,319	0.1995	0.2591
51	60	10	0.1099	12,310	0.1997	0.2603
55	60	50	0.1093	12,319	0.2002	0.2600
57	60	70	0.1085	12,321	0.2005	0.2601
53	60	30	0.1084	12,328	0.2006	0.2597
44	50	40	0.1147	12,357	0.2013	0.2594
42	50	20	0.1121	12,368	0.2035	0.2611
45	50	50	0.1117	12,376	0.2037	0.2606
50	50	100	0.1119	12,367	0.2037	0.2596
41	50	10	0.1112	12,380	0.2040	0.2628
48	50	80	0.1108	12,377	0.2045	0.2619
47	50	70	0.1094	12,377	0.2052	0.2606
49	50	90	0.1089	12,374	0.2059	0.2596
46	50	60	0.1086	12,373	0.2062	0.2611
43	50	30	0.1072	12,368	0.2071	0.2614
32	40	20	0.1125	12,424	0.2106	0.2627
35	40	50	0.1123	12,426	0.2107	0.2620
39	40	90	0.1116	12,419	0.2117	0.2615
31	40	10	0.1108	12,419	0.2123	0.2638
38	40	80	0.1108	12,418	0.2123	0.2621
34	40	40	0.1106	12,418	0.2125	0.2637
40	40	100	0.1101	12,417	0.2131	0.2621
36	40	60	0.1100	12,414	0.2132	0.2630
37	40	70	0.1092	12,424	0.2139	0.2621
33	40	30	0.1078	12,440	0.2144	0.2623
25	30	50	0.1121	12,483	0.2185	0.2655
23	30	30	0.1117	12,476	0.2189	0.2663
29	30	90	0.1110	12,480	0.2192	0.2653
28	30	80	0.1108	12,476	0.2192	0.2658
27	30	70	0.1107	12,474	0.2194	0.2660
22	30	20	0.1106	12,477	0.2195	0.2683
21	30	10	0.1098	12,489	0.2200	0.2694
30	30	100	0.1099	12,486	0.2200	0.2669
26	30	60	0.1083	12,488	0.2212	0.2674
24	30	40	0.1081	12,486	0.2213	0.2681
18	20	80	0.1122	12,566	0.2308	0.2706
14	20	40	0.1106	12,560	0.2318	0.2722
12	20	20	0.1106	12,554	0.2318	0.2739
20	20	100	0.1106	12,559	0.2318	0.2717

Continued on next page

Table 4.4 – *Continued from previous page*

ID	Cost	Cross	Sigma	NSV	Trer	Crer
19	20	90	0.1105	12,556	0.2319	0.2704
13	20	30	0.1102	12,562	0.2320	0.2715
15	20	50	0.1100	12,565	0.2323	0.2704
11	20	10	0.1092	12,560	0.2331	0.2718
16	20	60	0.1089	12,560	0.2335	0.2713
17	20	70	0.1085	12,565	0.2338	0.2716
10	10	100	0.1125	12,659	0.2492	0.2799
5	10	50	0.1121	12,656	0.2493	0.2805
6	10	60	0.1113	12,647	0.2497	0.2812
8	10	80	0.1107	12,657	0.2505	0.2818
4	10	40	0.1105	12,656	0.2506	0.2814
3	10	30	0.1101	12,658	0.2508	0.2817
1	10	10	0.1099	12,655	0.2509	0.2812
7	10	70	0.1095	12,649	0.2511	0.2817
2	10	20	0.1088	12,651	0.2515	0.2820
9	10	90	0.1087	12,652	0.2516	0.2813

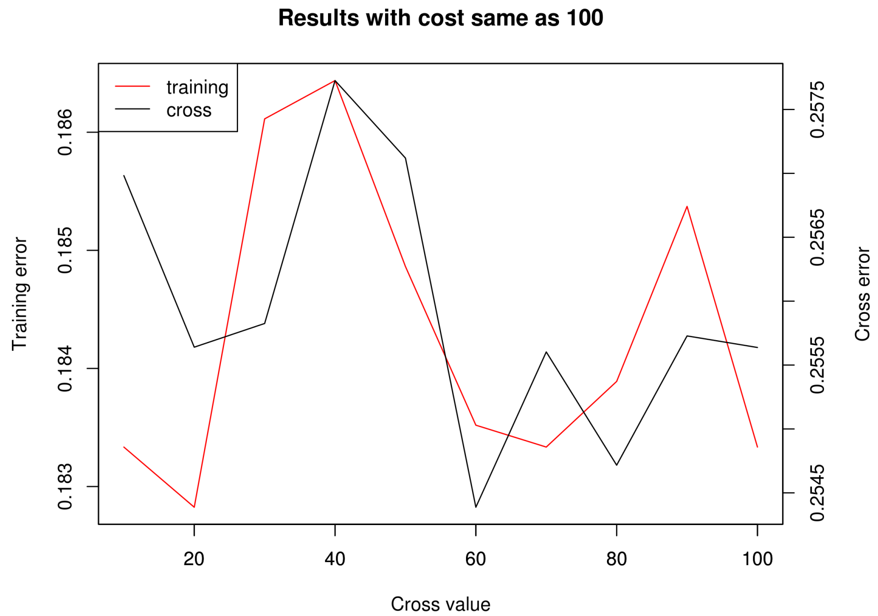


Figure 4.6: Results with cost same as 100 for Sendai

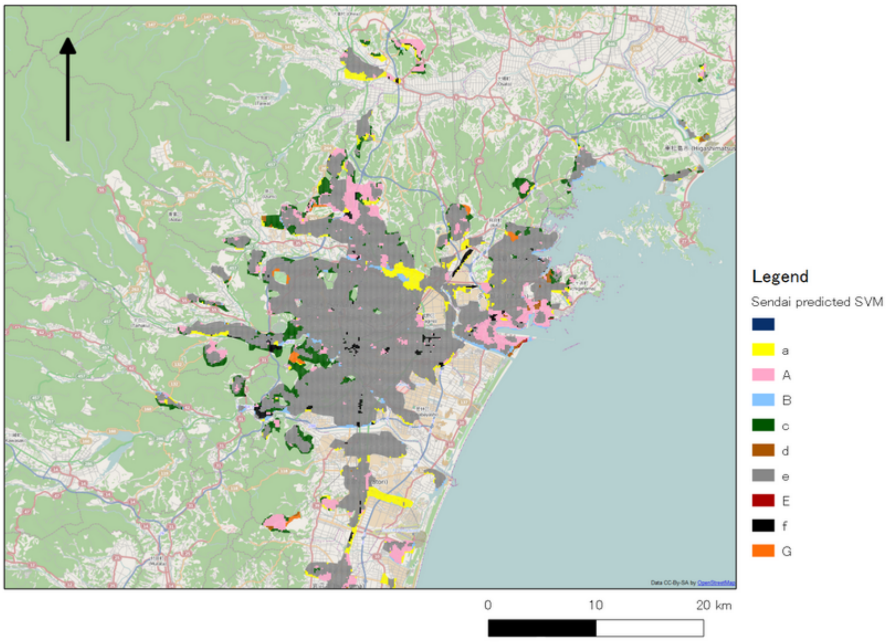


Figure 4.7: UPA of Sendai MtA predicted using SVM model

The best parameters for the SVM model are described as follows:

Cost value	100
Cross value	20
Sigma	0.12
Number of support vectors	12,209
Training error	0.18
Cost error	0.26
Kappa Index	0.7
Rand Index	0.8

Unlike Aomori MtA experiment, the training and cost error are bigger, however good Kappa and Rand indexes were calculated using this model. Predicted UPA of Sendai MtA is shown in Figure 4.7.

4.2.3 *UPA of Sapporo MtA*

The UPA of Sapporo MtA is the largest area for this analysis, for that reason, both the computational time and the memory resources are considerably high. Figure 4.8 shows the procedure to predict the land use factor using SVM.

The process to find the best model is as follows:

1. Create polygons of size 1, 3, 5, 10 and 20 km in the $mesh_i$ and intercept them with the socio-economic factors. Because the com-

putational process in this particular MtA was expensive, it was necessary to calculate the matrix Origin-Destination in a different way. Hence, a n-dimensional matrix was constructed in order to gain computational speed, reducing hardware resources and memory.

2. Save the information of overlaid objects depending on the polygon size in a list. Creating this list the amount of computational memory decreases, it can be verified using the size of the origin - destination (OD) matrix between the UPA size and all the socio-economic factors.

3. Calculate the closest socio-economic factors for the mesh$_i$.

4. Merging the Geographical information of the UPA with the database of distances.

5. Database cleansing process.

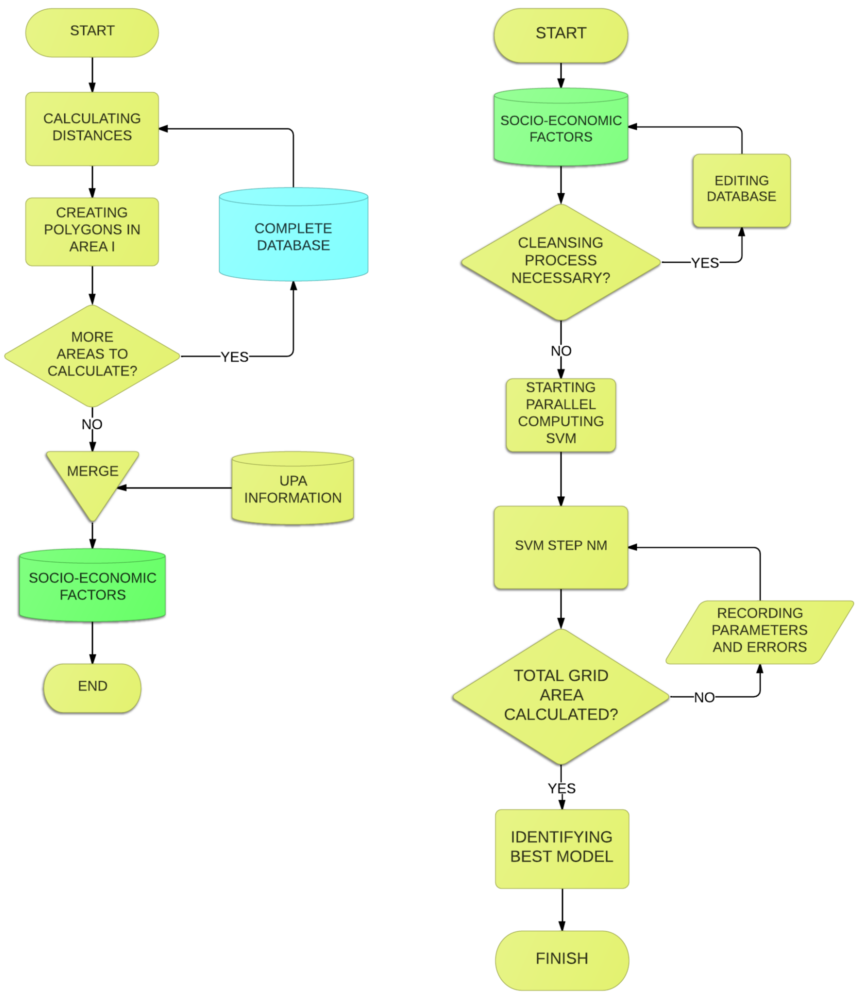


Figure 4.8: SVM procedure calculation for the UPA of Sapporo MtA

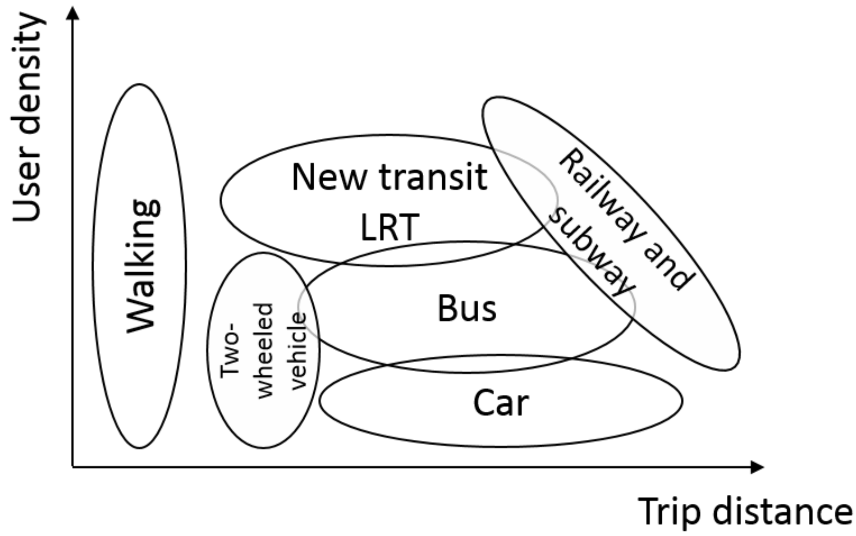


Figure 4.9: Adapted from "Comparison of the public transport system"[3]

6. Define total grid area.
7. Define tasks for each processor (Defining SVM's parameters).
8. Calculate SVM model using predefined parameters.
9. Once all the processors have finished, the identification of best model can be accomplished.

4.2.3.1 *Polygons radius*

The polygons radius were chosen according to the compact city definition that pursue a city with short distances between housing and socio-economic factors. Figure 4.9 shows the interactions between different transportation methods and user densities, LRT refers to *Light Rail Transit*.

4.2.3.2 *Walking*

The average human walking speed is about 5 km/h. However, it depends on different conditions, for this study, I considered their ages. For elderly pedestrians the average walking speed is between 4.51 - 4.75 km/h, and for younger individuals between 5.32 - 5.43 km/h. For this study I have considered 4.5 km/h as the average walking speed[4].

3 Comparison of the public transport system. https://wwwtb.mlit.go.jp/chubu/kikaku/chikousin/toshin2/toshin2_shiryo5.pdf

4 Walking. http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Walking.html, last access: 03/06/2014

Table 4.5: Traveling time by transportation methods (h)

Distance (km)	Time (h)				
	Walking	Bicycle	Bus	Train	Car
1	0.22	0.06	0.10	0.03	0.03
3	0.67	0.17	0.30	0.10	0.10
5	1.11	0.28	0.50	0.17	0.17
10	2.22	0.56	1.00	0.33	0.25
20	4.44	1.11	2.00	0.67	0.50
Farther distances	-	-	-	✓	✓

4.2.3.3 *Average bike riding speed*

The average bike riding speed in a city is between 17.7 - 19.3 km/h. As well as walking, the speed depends on different factors; however, I considered the average speed[5] as 18 km/h.

4.2.3.4 *Average public bus speed*

For public buses I considered an average speed of 10 km/h according to the MLIT[6].

4.2.3.5 *Average car speed*

The average car speed in urban areas is 40 km/h, on highways it is between 80 - 100 km, across the streets is 30 km/h and elsewhere between 50 - 60 km/h. For this study I considered the speed across the streets as 30 km/h and 40 km/h in the urban area[7].

4.2.3.6 *Average train speed*

There are different types of train, however I have taken as average speed inside the urban area[8] 30 km/h.

In order to describe why and how the intervals were defined a traveling time is calculated and shown in table 4.5. The cells highlighted show the most appropriate transportation system between the different distances.

5 Average bike riding speed. http://www.livestrong.com/article/413599-the-average-bike-riding-speed/, last access: 03/06/2014
6 Comparison of the public transportation system. https://wwwtb.mlit.go.jp/chubu/kikaku/chikousin/toshin2/toshin2_shiryo5.pdf, last access: 03/06/2014
7 Driving a car. http://www.japan-guide.com/e/e2022.html, last access: 03/06/2014
8 Comparison of the public transportation system

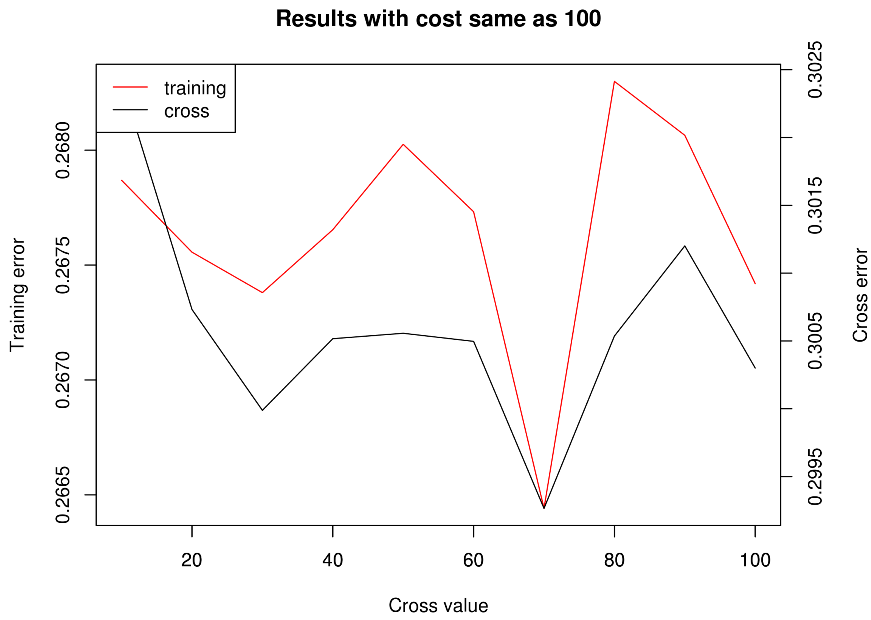


Figure 4.10: Results with cost same as 100 for Sapporo

The results for the SVM experiments in the UPA of Sapporo MtA are shown in Table 4.6. Same as the SVM experiments in the UPA of Aomori MtA, these results show that when the cost was 100 the training error was minimum (Figure 4.10).

Table 4.6: Sapporo SVM results

ID	Cost	Cross	Sigma	NSV	Trer	Crer
97	100	70	0.1801	30,732	0.2664	0.2993
93	100	30	0.1773	30,730	0.2674	0.3000
100	100	100	0.1772	30,735	0.2674	0.3003
92	100	20	0.1765	30,744	0.2676	0.3007
94	100	40	0.1760	30,750	0.2677	0.3005
96	100	60	0.1757	30,762	0.2677	0.3005
91	100	10	0.1752	30,750	0.2679	0.3024
95	100	50	0.1747	30,749	0.2680	0.3006
99	100	90	0.1743	30,766	0.2681	0.3012
98	100	80	0.1736	30,768	0.2683	0.3005
89	90	90	0.1802	30,789	0.2684	0.3005
81	90	10	0.1789	30,800	0.2687	0.3028
87	90	70	0.1784	30,799	0.2690	0.3011
83	90	30	0.1782	30,799	0.2691	0.3024
90	90	100	0.1764	30,808	0.2695	0.3007
85	90	50	0.1764	30,814	0.2695	0.3014
82	90	20	0.1756	30,820	0.2696	0.3015
86	90	60	0.1747	30,829	0.2698	0.3014
88	90	80	0.1748	30,815	0.2698	0.3011
84	90	40	0.1737	30,811	0.2703	0.3019
74	80	40	0.1795	30,850	0.2705	0.3014
73	80	30	0.1781	30,854	0.2709	0.3017
80	80	100	0.1775	30,862	0.2711	0.3021
78	80	80	0.1769	30,868	0.2713	0.3022
76	80	60	0.1764	30,870	0.2714	0.3015
75	80	50	0.1763	30,863	0.2715	0.3031
77	80	70	0.1756	30,871	0.2716	0.3023
71	80	10	0.1752	30,865	0.2718	0.3031
79	80	90	0.1750	30,866	0.2719	0.3024
72	80	20	0.1729	30,866	0.2726	0.3042
62	70	20	0.1768	30,892	0.2737	0.3051
67	70	70	0.1766	30,891	0.2737	0.3031
68	70	80	0.1760	30,911	0.2738	0.3036
65	70	50	0.1760	30,899	0.2738	0.3037
69	70	90	0.1757	30,903	0.2740	0.3035
64	70	40	0.1753	30,906	0.2741	0.3037
61	70	10	0.1751	30,907	0.2741	0.3055
66	70	60	0.1747	30,918	0.2742	0.3043
63	70	30	0.1747	30,913	0.2742	0.3042
70	70	100	0.1737	30,913	0.2743	0.3039
60	60	100	0.1780	30,979	0.2759	0.3043

Continued on next page

Table 4.6 – *Continued from previous page*

ID	Cost	Cross	Sigma	NSV	Trer	Crer
53	60	30	0.1779	30,980	0.2760	0.3051
57	60	70	0.1778	30,981	0.2760	0.3046
51	60	10	0.1772	30,981	0.2762	0.3056
59	60	90	0.1771	30,984	0.2762	0.3051
52	60	20	0.1769	30,984	0.2764	0.3058
56	60	60	0.1762	30,988	0.2767	0.3052
54	60	40	0.1756	31,000	0.2769	0.3048
55	60	50	0.1746	31,005	0.2772	0.3053
58	60	80	0.1730	31,010	0.2776	0.3057
48	50	80	0.1785	31,087	0.2792	0.3065
49	50	90	0.1778	31,086	0.2795	0.3072
42	50	20	0.1770	31,080	0.2796	0.3074
43	50	30	0.1773	31,097	0.2796	0.3074
45	50	50	0.1767	31,085	0.2796	0.3063
46	50	60	0.1764	31,090	0.2797	0.3068
47	50	70	0.1762	31,097	0.2797	0.3071
50	50	100	0.1747	31,084	0.2802	0.3070
41	50	10	0.1742	31,087	0.2805	0.3092
44	50	40	0.1726	31,094	0.2812	0.3081
36	40	60	0.1800	31,204	0.2836	0.3081
38	40	80	0.1791	31,203	0.2838	0.3088
32	40	20	0.1783	31,205	0.2839	0.3099
31	40	10	0.1777	31,199	0.2840	0.3114
39	40	90	0.1767	31,211	0.2844	0.3094
35	40	50	0.1755	31,179	0.2845	0.3097
37	40	70	0.1748	31,187	0.2849	0.3100
40	40	100	0.1749	31,183	0.2849	0.3095
33	40	30	0.1729	31,179	0.2854	0.3108
34	40	40	0.1696	31,230	0.2865	0.3109
22	30	20	0.1783	31,301	0.2894	0.3123
23	30	30	0.1781	31,309	0.2895	0.3131
29	30	90	0.1776	31,318	0.2896	0.3122
21	30	10	0.1766	31,322	0.2897	0.3133
24	30	40	0.1751	31,355	0.2901	0.3124
28	30	80	0.1751	31,359	0.2901	0.3124
27	30	70	0.1749	31,343	0.2901	0.3126
25	30	50	0.1737	31,355	0.2902	0.3131
30	30	100	0.1730	31,356	0.2904	0.3134
26	30	60	0.1725	31,373	0.2904	0.3135
16	20	60	0.1785	31,588	0.2971	0.3176
20	20	100	0.1782	31,585	0.2972	0.3177
14	20	40	0.1770	31,584	0.2977	0.3176
13	20	30	0.1766	31,582	0.2977	0.3184

Continued on next page

Table 4.6 – *Continued from previous page*

ID	Cost	Cross	Sigma	NSV	Trer	Crer
15	20	50	0.1763	31,577	0.2978	0.3176
17	20	70	0.1763	31,576	0.2978	0.3177
18	20	80	0.1761	31,578	0.2978	0.3172
19	20	90	0.1751	31,572	0.2981	0.3177
11	20	10	0.1736	31,571	0.2987	0.3194
12	20	20	0.1727	31,567	0.2988	0.3198
8	10	80	0.1783	31,885	0.3094	0.3263
2	10	20	0.1778	31,881	0.3096	0.3261
5	10	50	0.1770	31,884	0.3097	0.3265
1	10	10	0.1759	31,891	0.3100	0.3277
6	10	60	0.1758	31,886	0.3100	0.3265
10	10	100	0.1757	31,900	0.3100	0.3261
3	10	30	0.1753	31,898	0.3101	0.3271
7	10	70	0.1748	31,906	0.3102	0.3260
4	10	40	0.1744	31,882	0.3102	0.3273
9	10	90	0.1741	31,887	0.3103	0.3268

The best parameters for the SVM model are:

ID	97
Cost	100
Cross value	70
Sigma	0.18
Number of support vectors	30,732
Training error	0.26
Cross error	0.29
Kappa value	0.59
Rand index	0.76

The results present a training error of 26.6%, for that reason further results are necessary to reduce the error. Using these parameters the contingency table is shown in Table 4.7. The predicted UPA of the MtA is shown in the Figure 4.11.

Further experiments using parallel computing were developed to reduce the training error. The minimum training error was found after 13.9 days. Doing these experiments, it was possible to reduce the training error by 5%. The results are shown in Table 4.8. Finally, I used the best model from the extended experiments shown as follows in order to predict the socio-economic factors in this MtA:

Table 4.7: Contingency table for the UPA of Sapporo MtA

	a	c	d	e	A	B	E	f	G
a	1,323	199	143	100	100	27	0	17	3
c	275	3,631	396	203	256	95	3	83	1
d	96	298	3,046	82	254	136	36	92	0
e	800	986	510	23,731	4,134	476	14	1,375	4
A	169	336	280	943	4,893	184	8	228	11
B	20	17	35	30	61	623	6	12	0
E	0	0	2	0	1	0	30	0	0
f	8	17	18	13	18	0	0	180	0
G	0	2	0	1	5	1	0	1	44

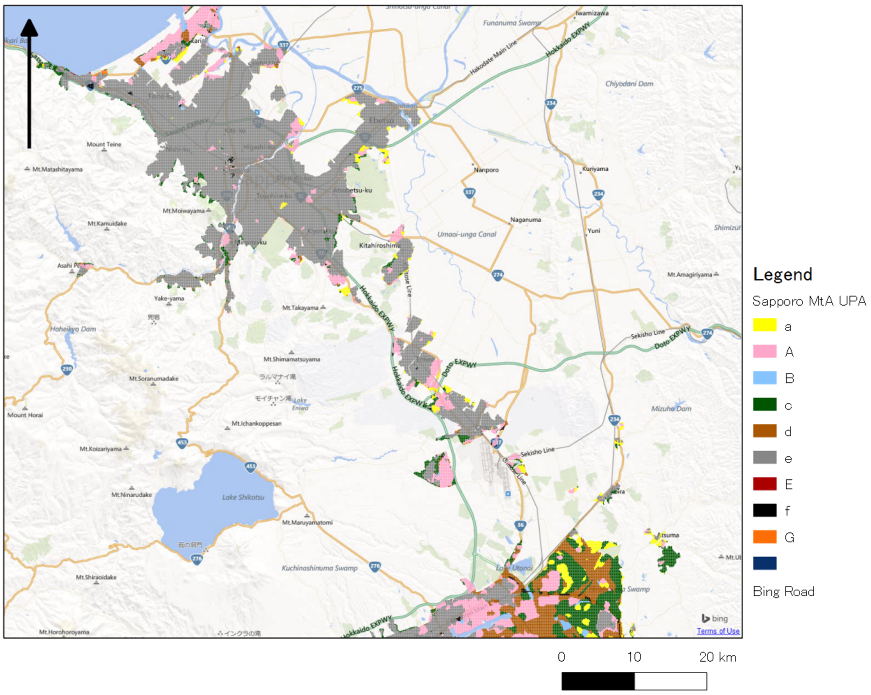


Figure 4.11: UPA of Sapporo MtA predicted using SVM model

Cost	1,000
Cross value	20
Sigma	0.18
Number of support vectors	29,536
Training error	0.21
Cross error	0.28

A new proposal for the UPA of Sapporo MtA calculated by using SVM is shown in Figure 4.12 .

Table 4.8: Extended experiments for the UPA of Sapporo MtA

Cost	Cross	Sigma	NSV	Trer	Crer
1,000	20	0.1785	29,536	0.2162	0.2859
1,000	10	0.1768	29,551	0.2168	0.2869
1,000	50	0.1766	29,545	0.2169	0.2849
1,000	30	0.1759	29,542	0.2172	0.2848
1,000	40	0.1727	29,538	0.2188	0.2849
800	20	0.1783	29,611	0.2216	0.2851
800	50	0.1771	29,607	0.2221	0.286
800	10	0.1769	29,621	0.2222	0.2875
800	40	0.1753	29,626	0.2226	0.286
800	30	0.1746	29,623	0.223	0.2864
600	40	0.179	29,804	0.2282	0.2871
600	20	0.1776	29,813	0.2287	0.2892
600	30	0.1766	29,796	0.2292	0.2885
600	50	0.1761	29,785	0.2294	0.2878
600	10	0.1754	29,792	0.2296	0.2911
400	20	0.1781	29,978	0.2372	0.2907
400	10	0.1766	29,981	0.2379	0.2926
400	50	0.1762	29,990	0.2381	0.2899
400	40	0.1757	29,979	0.2383	0.2895
400	30	0.1742	29,993	0.2388	0.2909
200	40	0.1792	30,291	0.2526	0.2942
200	10	0.1784	30,283	0.2529	0.2974
200	50	0.1772	30,289	0.2532	0.2957
200	20	0.1762	30,301	0.2535	0.2964
200	30	0.172	30,319	0.255	0.2958

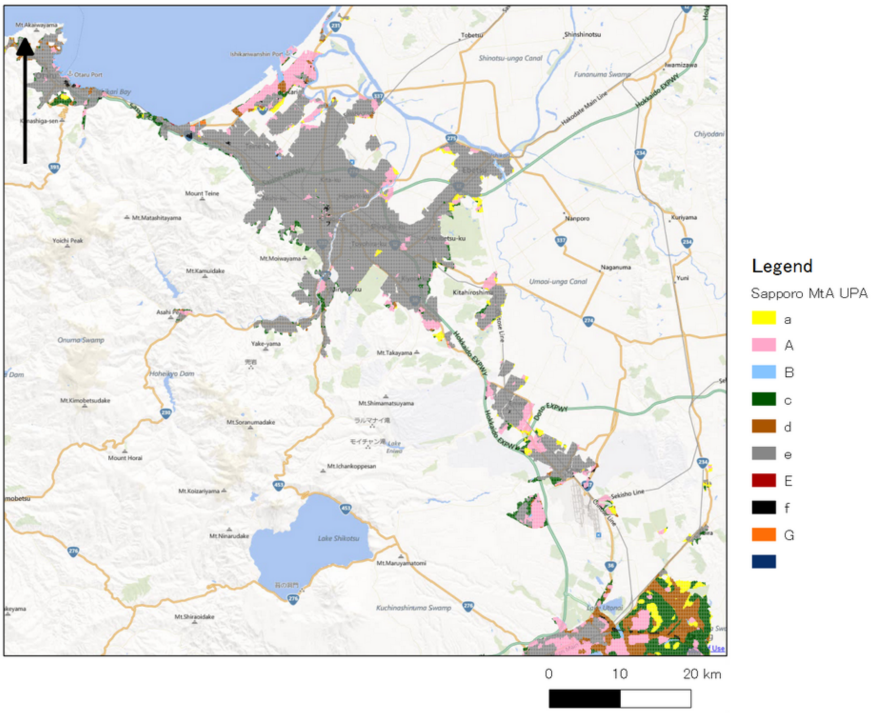


Figure 4.12: UPA of Sapporo MtA - (SVM alternative)

4.2.4 *Socio-economic factors indicators*

The indicators related to socio-economic factors are defined as follows:

4.2.4.1 *Health:*

The indicator is calculated as the number of health facilities over population in the designated area. This ratio can be adjusted per 1,000 people. This indicator can also be referenced in the WHO Health System Strengthening (HSS) Handbook[9].

4.2.4.2 *Trains stations:*

The indicator is defined as the number of train stations per km^2[19].

4.2.4.3 *Bus stops:*

As well as the previous indicator, this indicator is defined as the number of bus stops per km^2.

4.2.4.4 *Convenience stores:*

The indicator is defined as the number of convenience stores per 10,000 person[95][23].

4.2.4.5 *Parks:*

The indicator is calculated by the number of parks per capita; it can be adjusted to per 1,000 person[94].

4.2.4.6 *Public facilities:*

It is defined as the number of public facilities per capita; it also can be adjusted per 1,000 person.

4.2.4.7 *Supermarkets:*

The indicator is defined as the number of supermarkets per capita; however it can be adjusted per 10,000 person[10][68][27]

In Table 4.9, the calculation method of socio-economic factors (SEF) indicators is introduced. In some cases a multiplier is added according to the different references. In Figure 4.13, it is evident that for all the MtA the train stations are the socio-economic factors with less

9 Number and distribution of health facilities per 10,000 population. http://www.cpc.unc.edu/measure/prh/rh_indicators/crosscutting/hss/number-and-distribution-of-health-facilities-per, last access: 17/06/2014

10 Food For Every Child: The Need for More Supermarkets in Massachusetts. http://www.mphaweb.org/documents/FoodforEveryChild_Massachusetts.pdf, last access: 17/06/2014

Table 4.9: Socio-economic factors indicators

SEF	Multiplier	Indicator	Aomori	Sendai	Sapporo
Health	1,000	$\frac{facilities}{population}$	19.102	9.936	11.716
Train stations	1	$\frac{train.st}{km^2}$	0.001	0.001	0.006
Bus stops	1	$\frac{bus.st}{km^2}$	75.526	3.773	12.923
Convenience stores	10,000	$\frac{conv.st}{population}$	5.008	1.495	1.434
Parks	1,000	$\frac{parks}{population}$	0.946	1.146	1.360
Public facilities	10,000	$\frac{facilities}{population}$	40.362	11.467	11.408
Supermarkets	10,000	$\frac{supermarkets}{population}$	1.067	0.219	0.459

impact in the area; followed by supermarkets. The concentration of bus stops in the UPA of Aomori MtA is more than 6 times the UPA of Sapporo MtA, and more than 20 the UPA of Sendai MtA. It suggest that this transportation system should be improved in the middle and large MtAs. There are more convenience stores per population in the UPA of Aomori MtA; nevertheless, for Sendai and Sapporo MtAs the indicator is similar. The concentration of Parks is greater in the UPA of Sapporo MtA than the UPAs of Sendai and Aomori MtAs respectively. There are more than 3.5 times of Public facilities in the UPA of Aomori MtA than in the other MtAs. In Sendai and Sapporo MtAs the indicator is similar, same as 11.4 public facilities per 10,000 persons. Finally, the UPA of Aomori MtA has a concentration supermarkets higher than in the UPAs of Sapporo and Sendai MtAs respectively, however the number of this socio-economic factor is close to 1 per 10,000 persons in every MtA.

The socio-economic factors indicators show that their concentration is higher in the UPA of Aomori MtA than the UPAs of Sapporo and Sendai MtAs.

In Figure 4.14 a correspondence analysis is developed in order to estimate the degree of correlation through proximity and distance. The socio-economic factors such as supermarkets, convenience stores and public facilities are much more related to the UPAs of Sapporo and Aomori MtAs. Bus stops are closer to Aomori MtA while train stations are not related with the MtA. Medical Institutions are distinguished by the UPAs of Sapporo and Sendai MtAs. Another aspect is that the UPA of Aomori MtA lies on the other side of Sendai MtA, this shows the distinctiveness of Aomori MtA in the group.

By Calculating the different indicators and through correspondence analysis it was clarified that Aomori MtA presents more concentration of socio-economic factors in its UPA, followed by the UPAs of Sapporo and Sendai MtAs respectively.

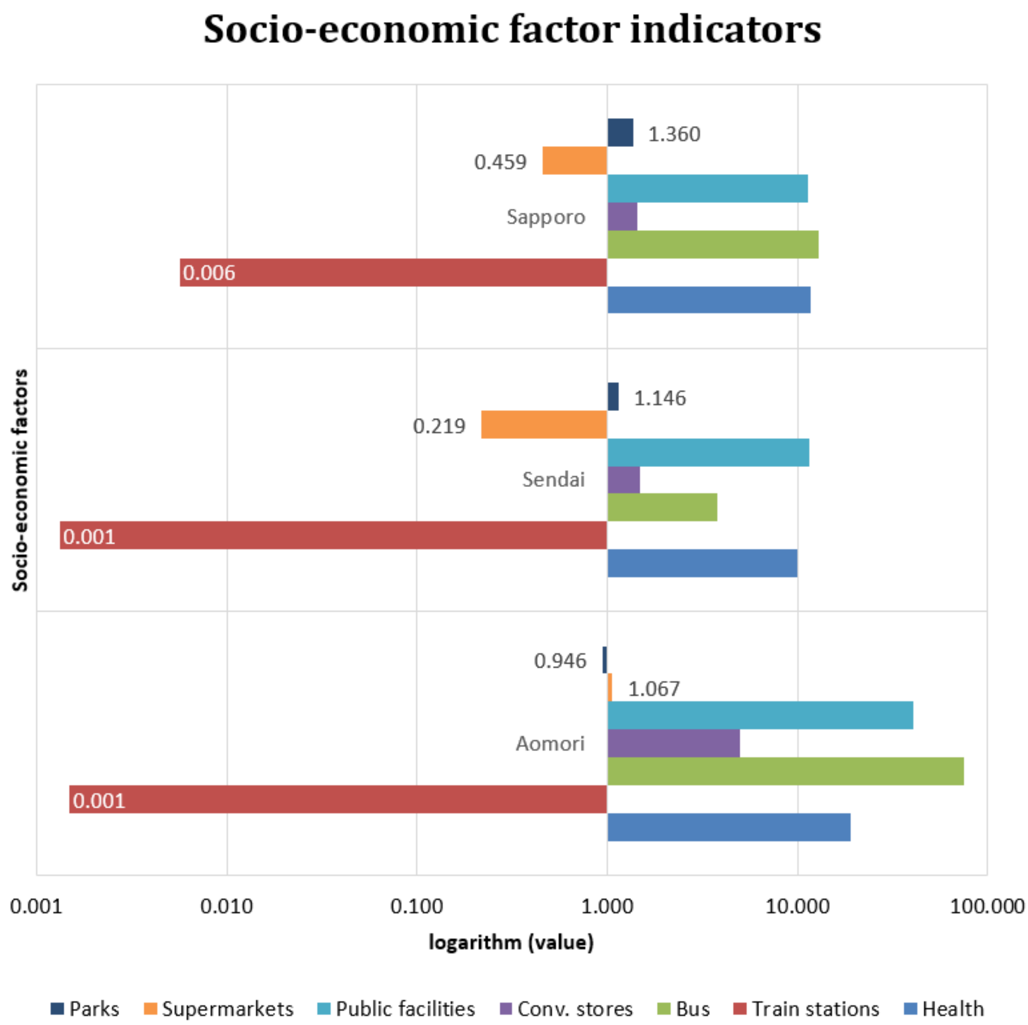


Figure 4.13: Socio-economic factors indicators

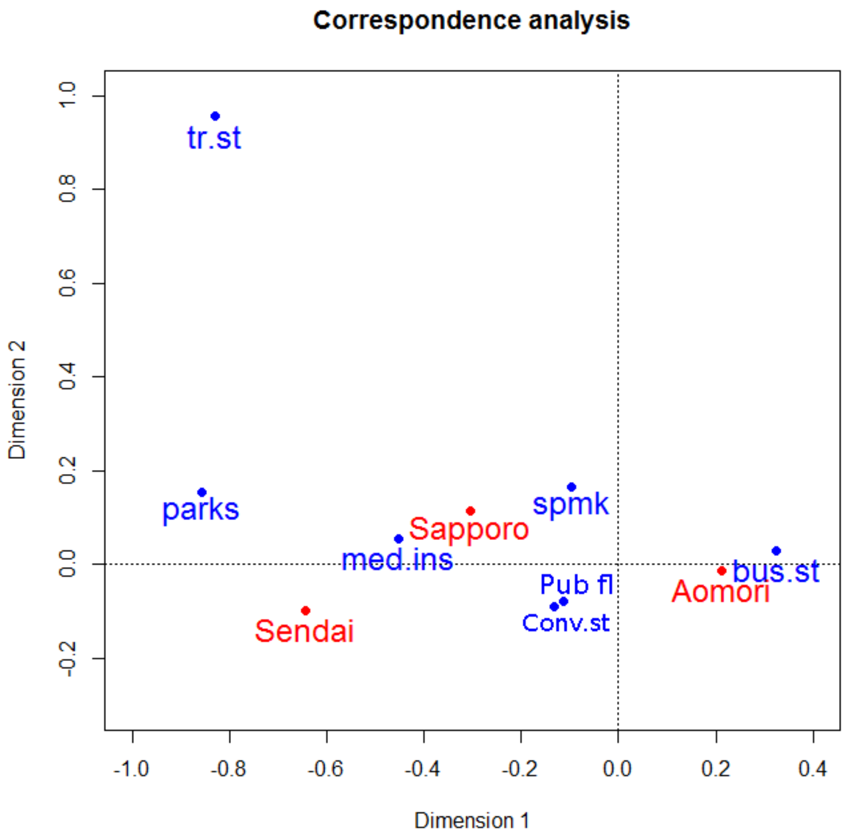


Figure 4.14: Correspondence analysis of socio-economic factors indicators

5 INTEGRATED EVALUATION

In this chapter parameters and results from previous models are gathered to make possible an integrated evaluation method. This chapter will explain first of all the methodology, highlighting the procedure for the evaluation method; it will be focused on the residential area in the UPA of Aomori MtA, because the goal of this area is to promote residential, commercial and industrial areas. Pattern recognition related to the relationship of land use and socio-economic factors will be addressed.

Thermodynamic and chemical concepts will be applied to construct a chemical compound by calculating relations between land use and socio-economic factors. Finally, an equation defining the evaluation of the compact city model is elaborated; through this equation, it will be possible to describe the current situation of all the MtAs and how far they are from their own goals.

5.1 METHODOLOGY

The first part of these experiments will be to analyze and create a chemical molecule that describe Aomori MtA's compact city model. In order to create the molecules the next process is necessary:

1. Loading the database containing the information of the socio-economic factors with the one of land use.

2. The question here is: Is the database in the Area of Interest (AOI)?, in order to identify if the data lies in this area it is necessary to overlay a polygon that contains the information of the AOI. In chapter 3, all the residual kriging models for population density have been shown in an interval composed of 5 sets. From that point of view a polygon is built taking the last 2 sets of each interval. In this way the relation between population and land use can be seen.

3. If all the databases do not lie in the AOI, it is necessary to overlay the polygon before mentioned, and create two databases containing the information that lies and another one with the data that doesn't lie in the AOI.

4. Using the information that lies in the AOI, it is necessary to extract each type of land use to build an appropriate chemical compound that describe the relation between population density, land use and socio-economic factors.

5. Extract the information about the land use (A) and socio-economic factors (B) and merged into a single database (C). Later, change the coordinate system into Universal Transverse Mercator (UTM).
6. Create marks for the previously mentioned database in order to identify each point.
7. Calculate nearest neighbors between data.
8. Calculate Lennard-Jones interaction and Strauss process results.
9. Calculate potential values between interactions.
10. Display compounds.

The flow diagram of the process is shown in Figure 5.1.

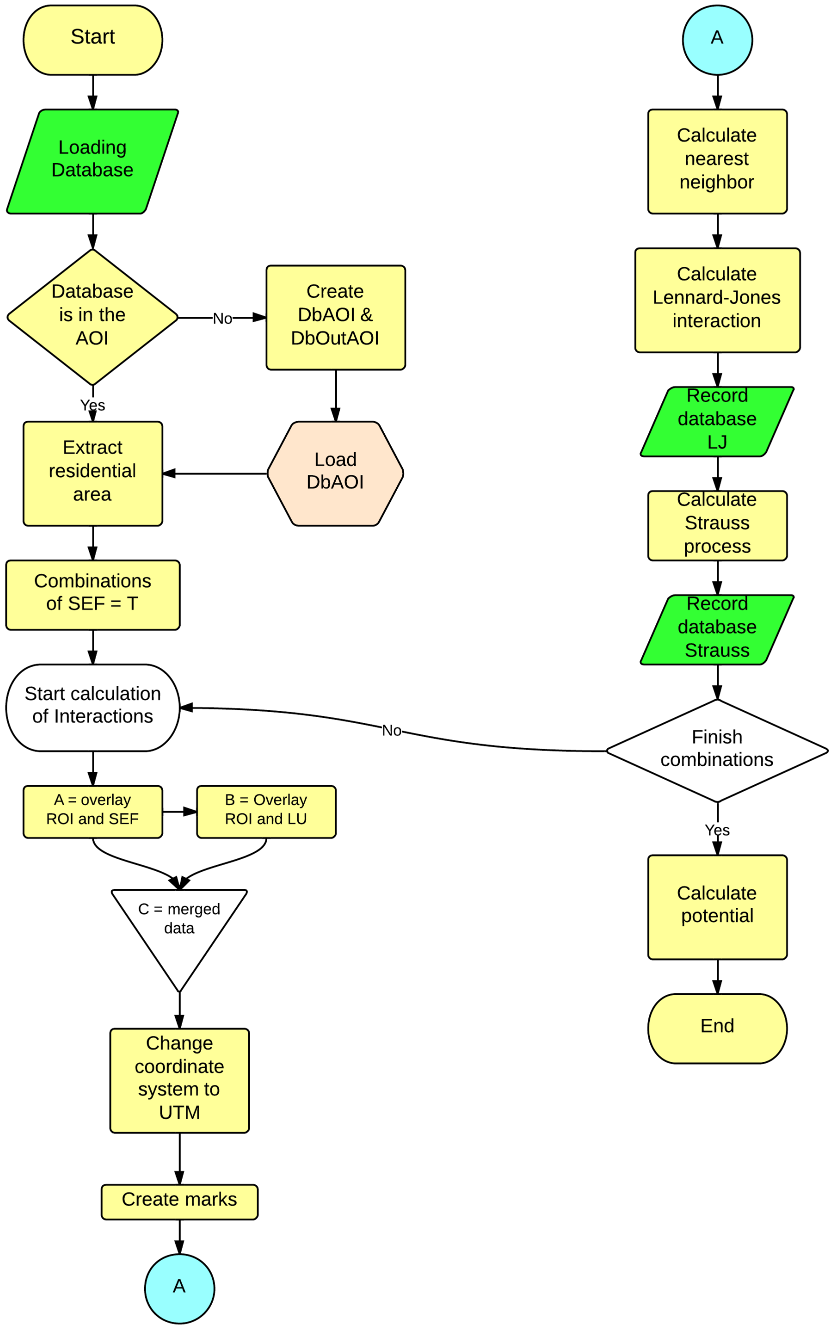


Figure 5.1: Calculating best chemical compound

5.2 INPUT DATA

Aomori MtA's compact city model is the study target, and it is required to extrapolate this model to other MtAs. For that reason, it is necessary to extract internal phenomena inside the Metropolitan urban core.

Table 5.1: Expected number of facilities per distance

	(0,1]	(1,3]	(3,5]	(5,20]
Bus stops	15.99	90.94	107.55	199.51
Convenience stores	4.26	23.66	23.94	31.13
Medical institutions	23.09	119.42	113.89	123.61
Other institutions	3.67	19.45	19.85	29.03
Parks	6.23	32.03	32.92	47.83
Price	4.61	23.96	24.29	36.15
Public facilities	23.85	130.16	138.30	202.69
Supermarkets	0.63	3.25	3.27	4.85
Train stations	0.48	2.37	2.47	6.68

5.2.1 *Distances to facilities*

According to AUC calculated in a previous chapter, it was possible to identify which are the most important socio-economic factors that affect the residential area.

The first part was to identify the expected number of facilities at different distance intervals measured in kilometers. Table 5.1 shows that there are almost 16 bus stops, 4 convenience stores, 23 medical institutions, 4 other institutions, 6 parks, 5 prices, 24 public facilities, 1 supermarket and 1 train station in an interval between 0 to 1 km.

In Figure 5.2 it is possible to identify that the distribution for number of bus stops in the interval between [0,1] has a positive skew; and expected value of this socio-economic factor is between 15 and 16 bus stops. Figure 5.3 shows in the same interval a reduction of the frequency, in an interval between 0 and 1 km it is possible to find 4 convenience stores around each mesh area. In Figure 5.4, in the interval between 0 to 1 km, the number of medical institutions shows a rapid decay. Between 0 and 3 km it is possible to find more than 130 medical institutions, however the density is 3 times smaller than the first interval. As well as medical institutions, Figure 5.5 presents a similar behavior, it is possible to find between 1 and 5 institutions in the first interval. Related to parks, Figure 5.6 shows that for most of the residents there are between 1 and 4 parks in this interval, this is good prevention policy in the case of natural disasters. However, it is necessary for all the residents to know where is the primary evacuation site in a possible event. The Figure 5.9 shows that there is a supermarket nearby between 0 and 1 km. Nevertheless, between 0 and 3 km it is possible to find at least 4 supermarkets around each mesh area. Finally, Figure 5.10 shows in the first interval that there is 1 train station for the residents in the UPA of this MtA.

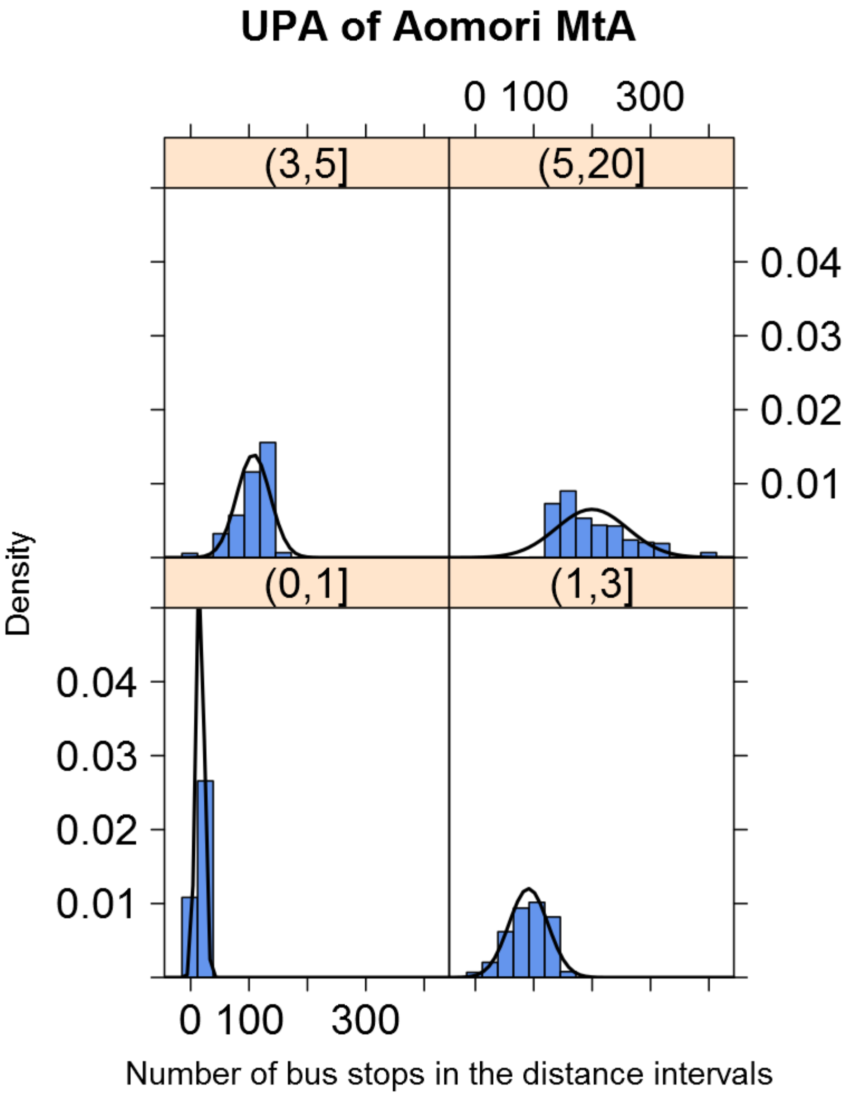


Figure 5.2: Number of bus stops per interval in the UPA of Aomori MtA

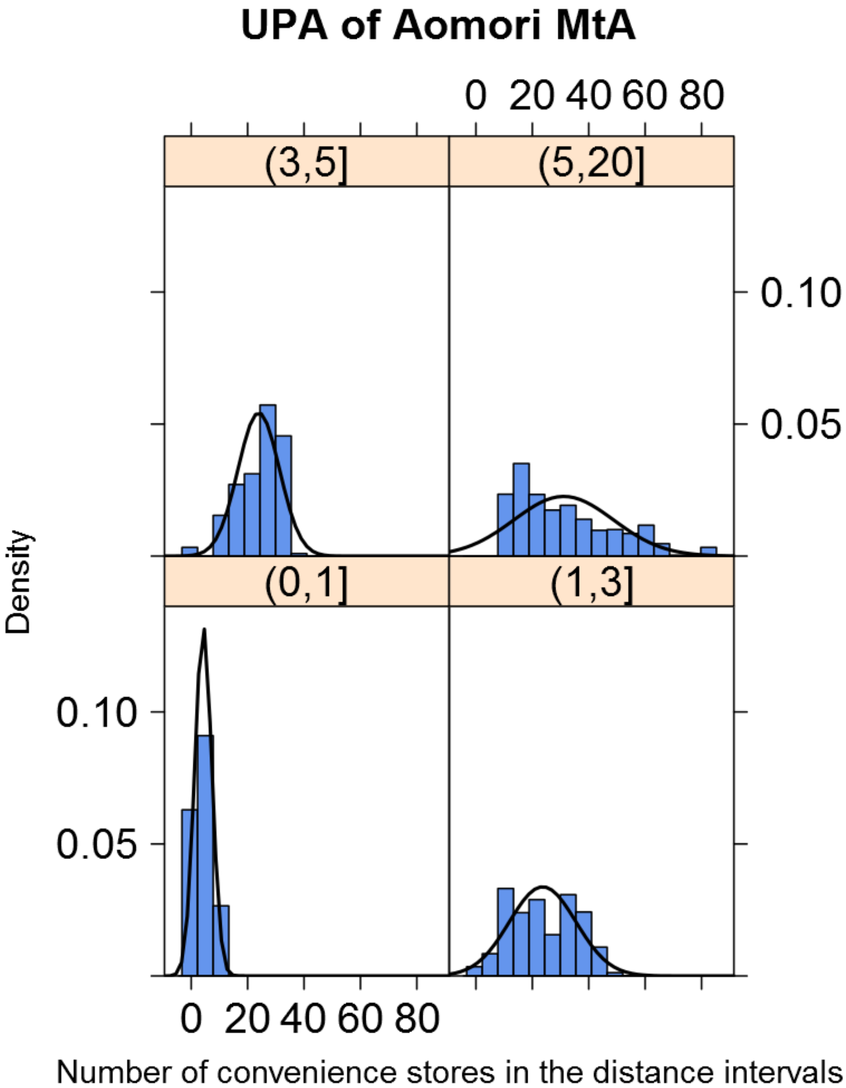


Figure 5.3: Number of convenience stores per distance interval in the UPA of Aomori MtA

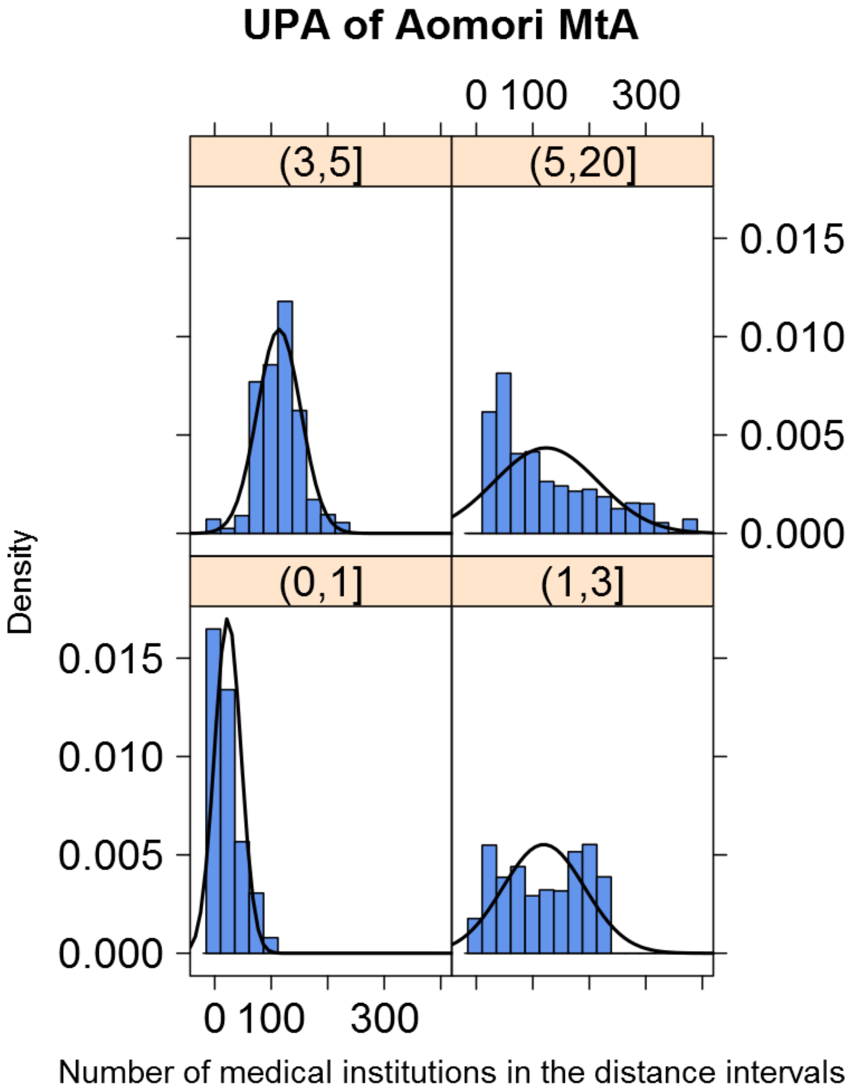


Figure 5.4: Number of medical institutions per distance interval in the UPA of Aomori MtA

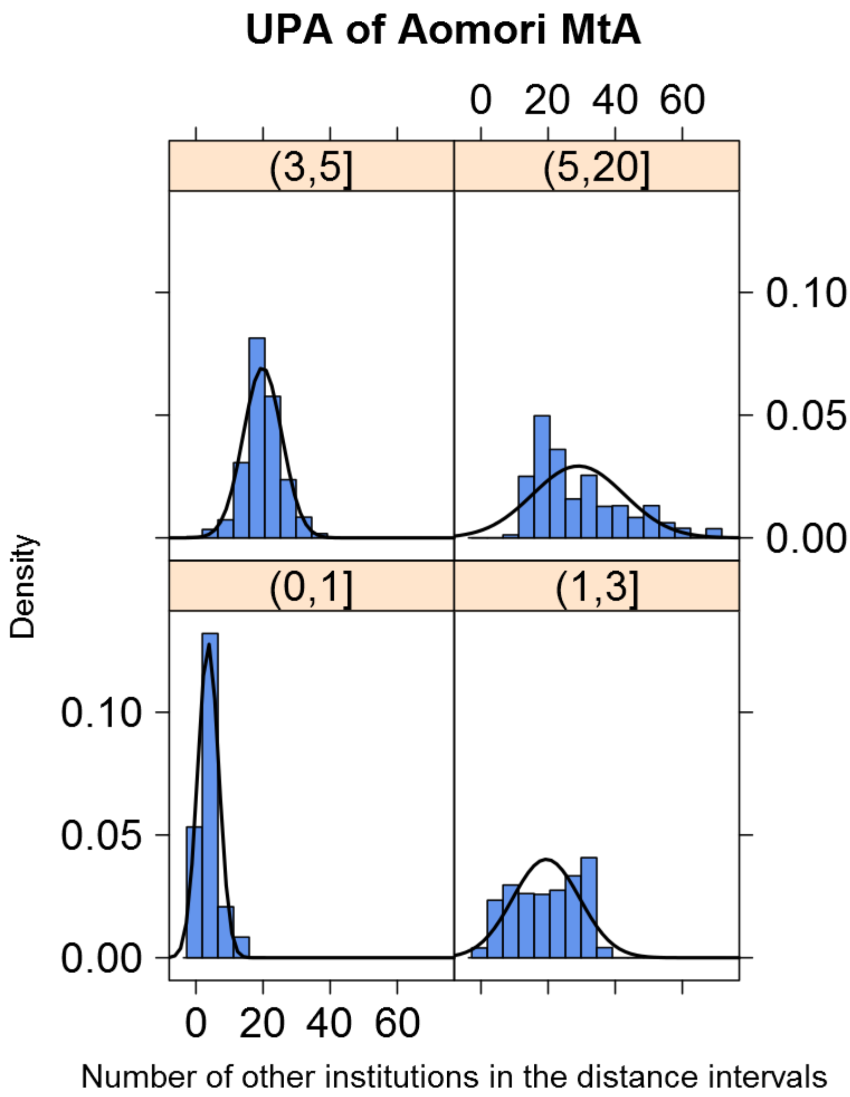


Figure 5.5: Number of other institutions per distance interval in the UPA of Aomori MtA

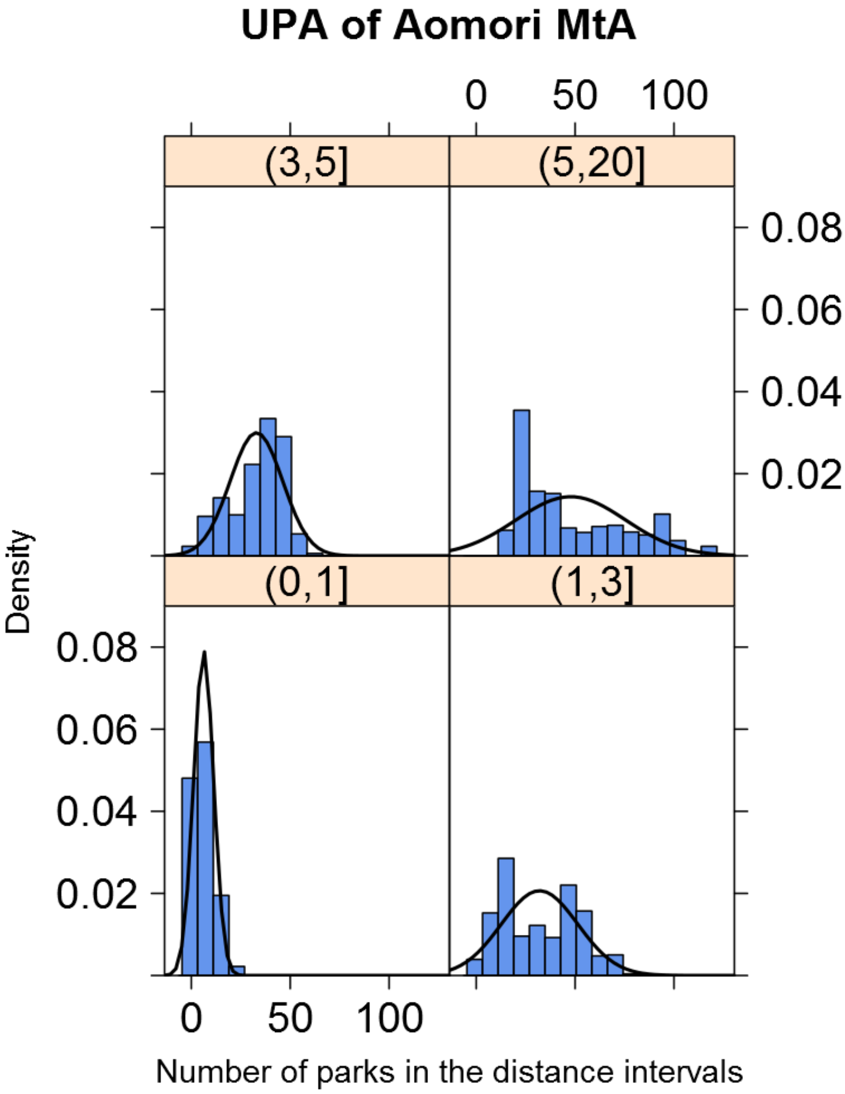


Figure 5.6: Number of parks per distance interval in the UPA of Aomori MtA

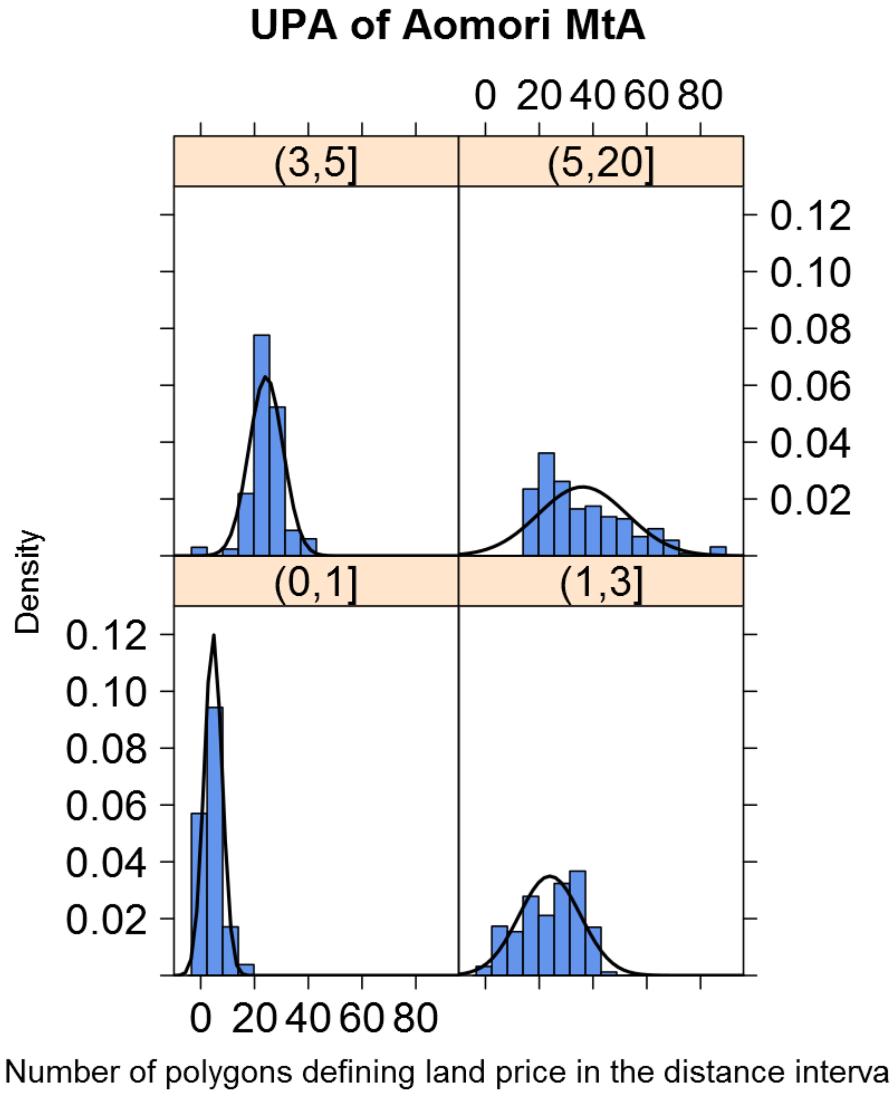


Figure 5.7: Number of polygons defining price per distance interval in the UPA of Aomori MtA

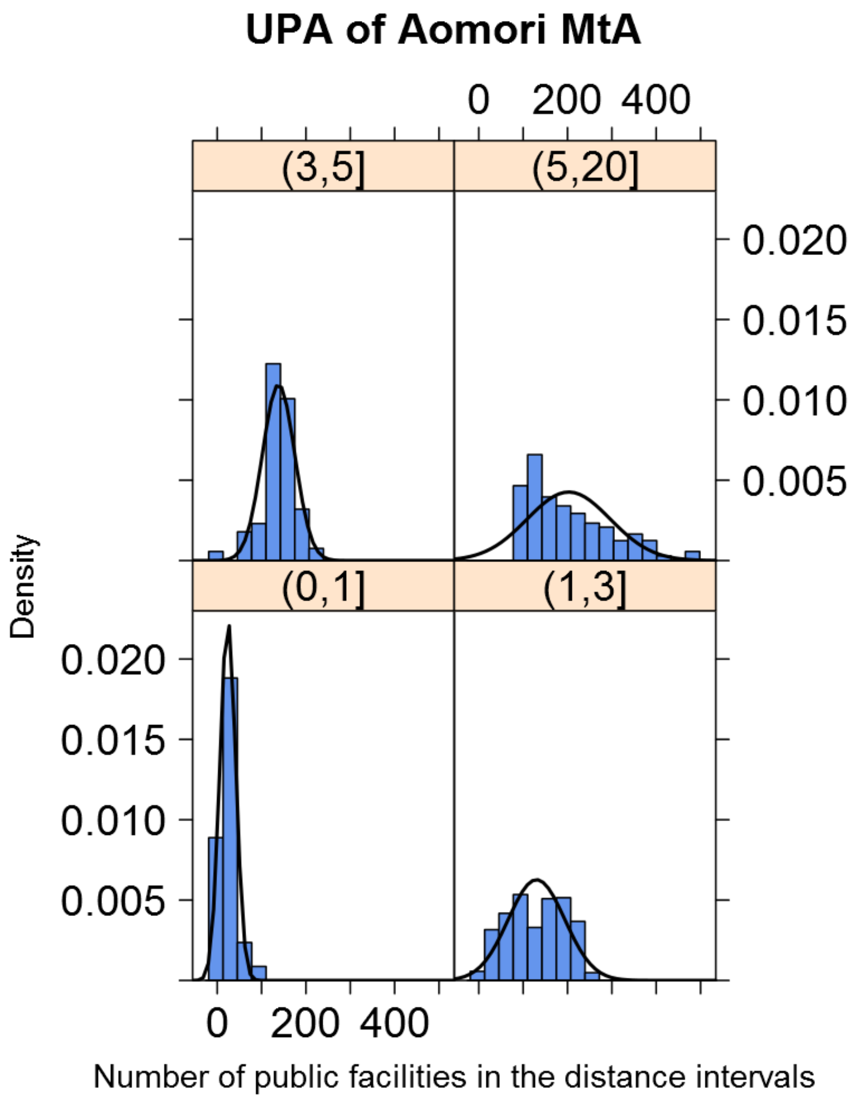


Figure 5.8: Number of public facilities per distance interval in the UPA of Aomori MtA

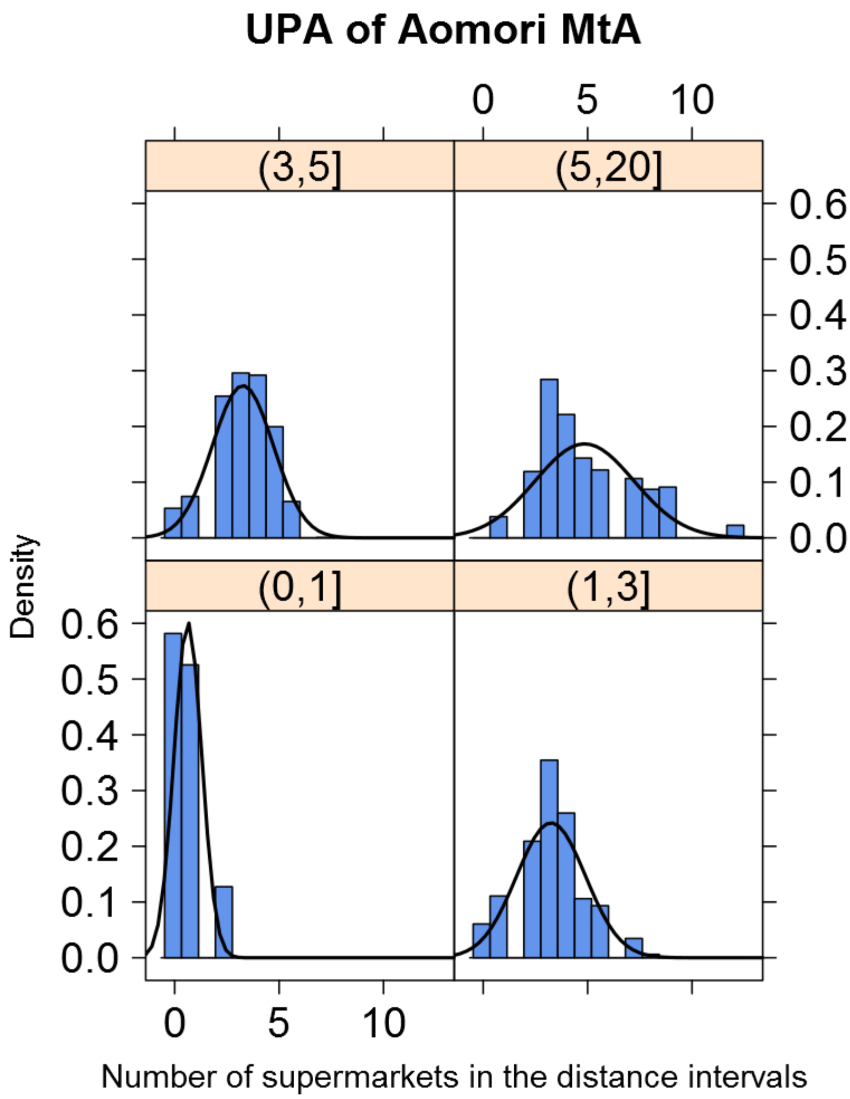


Figure 5.9: Number of supermarkets per distance interval in the UPA of Aomori MtA

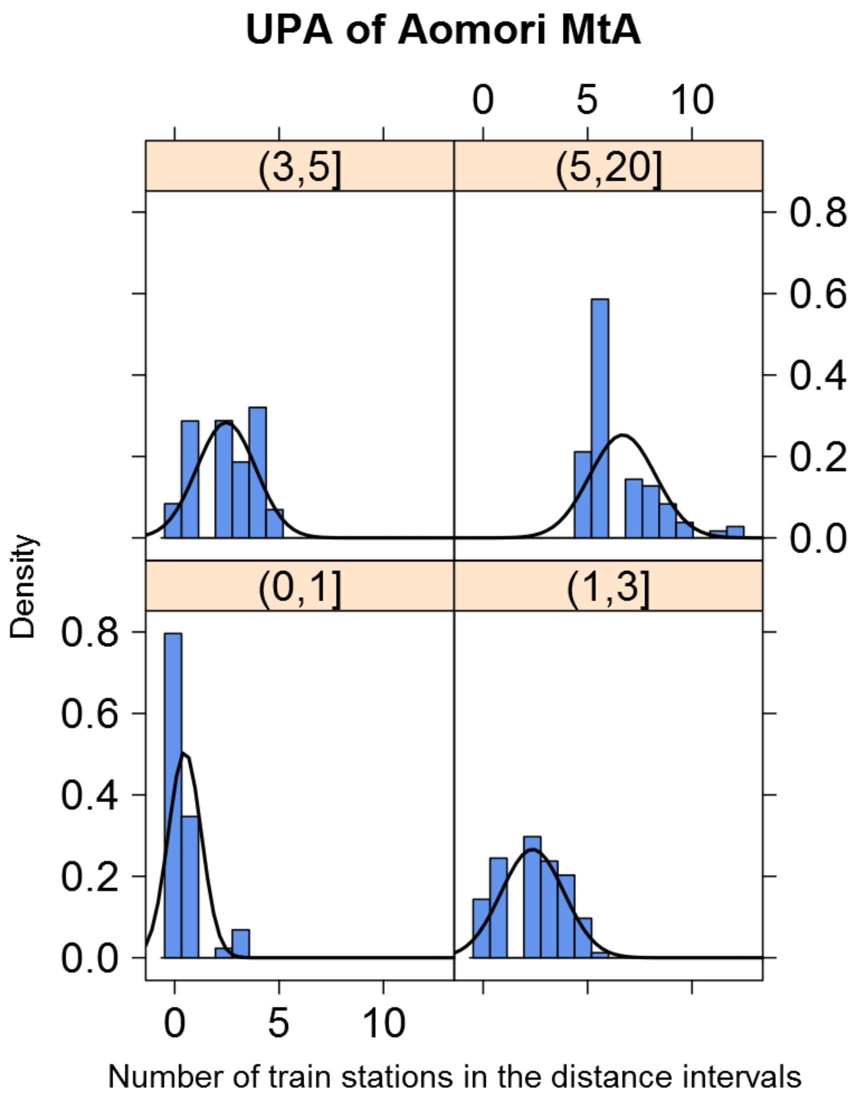


Figure 5.10: Number of train stations per distance interval in the UPA of Aomori MtA

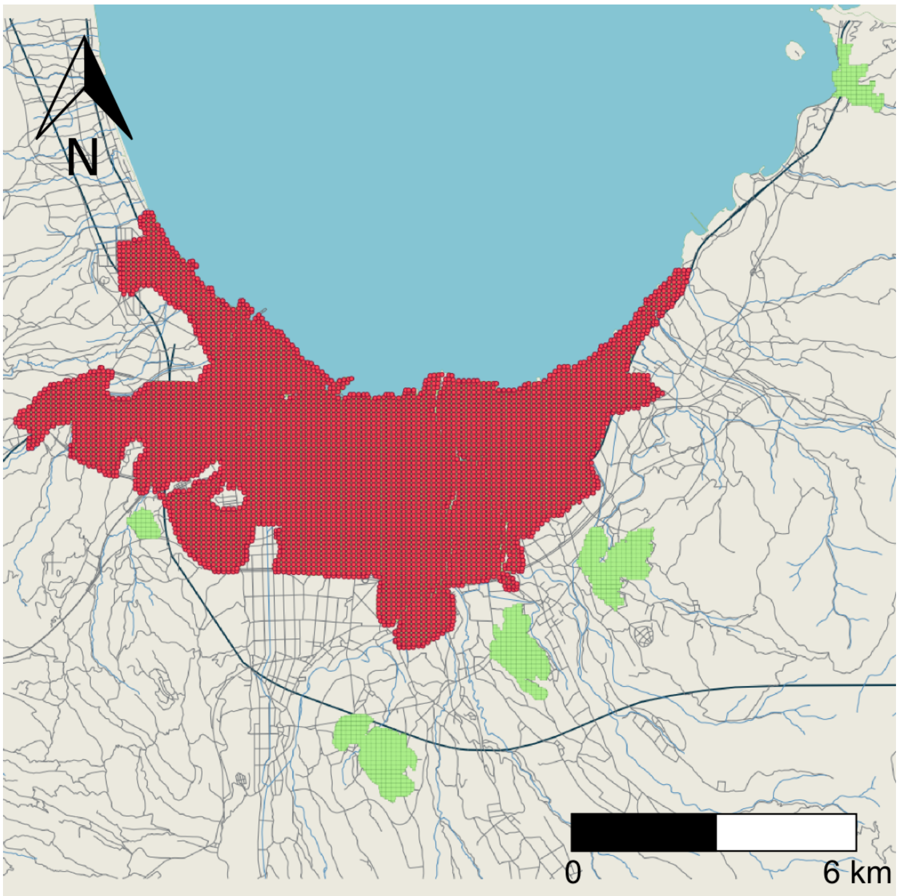


Figure 5.11: Overlaying land use data with compact city shape

5.2.2 *Residential area - special case*

The urban core of the UPA in Aomori MtA presents a total of 5,261 ha shown in red (see Figure 5.11). It is evident how the local government pursues one of the most important statements of the compact city model: Define properly the boundaries of the compact city.

This area corresponds to 87.2% of the total UPA. In Figure 5.12, the AUC for the residential area is shown. According to the UPA definition, this area should be promoted in the residential area, for that reason in this subsection I will focus on this particular land use type.

The socio-economic factors affect the residential area, for instance: convenience stores, medical institutions, public facilities, other institutions, distance to the polygon that defines price, supermarkets, bus stops and train stations.

Along the UPA urban core of this MtA, the residential area remains the largest land use type with 67.7% as shown in Table 5.2. This per-

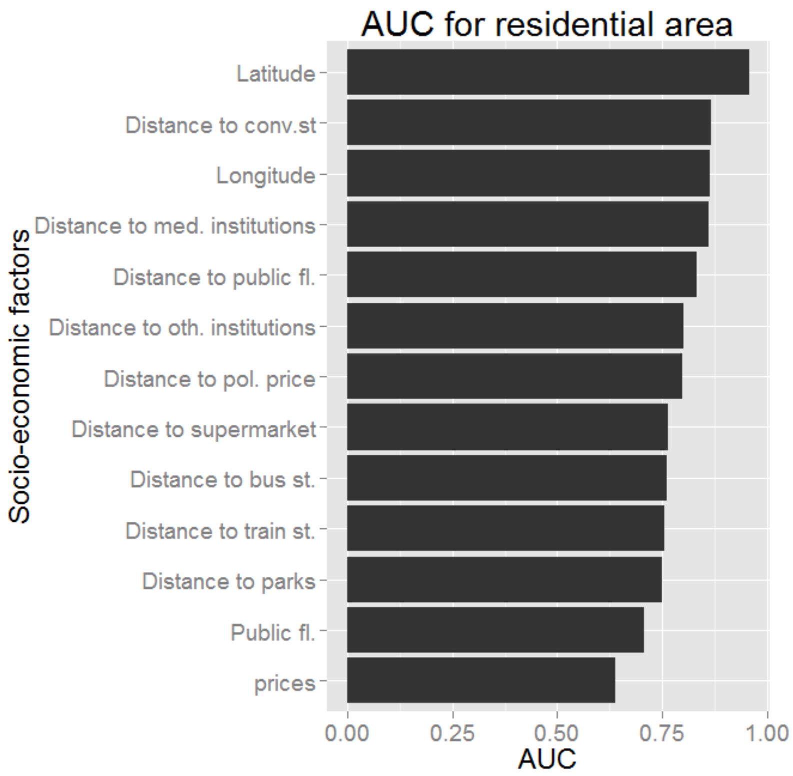


Figure 5.12: AUC for residential area

Table 5.2: Land use percentage in the UPA's core of Aomori's MtA

Land use type	Percentage (%)
Rice fields	5.056
Other agricultural land	1.407
Forest	1.787
Waste land	0.589
Residential area and buildings	67.687
Roads	5.436
Other sites	13.381
Rivers and lakes	1.901
Beach	0.038
Ocean	2.718

centage corresponds to 3,561 ha or 59.0% of the total UPA.

Later, the set of residential area located inside the UPA was extracted. This was useful to identify the number of vectors that describe the different relations between this type of land use and socio-

Table 5.3: Example of a matrix containing distance to socio-economic factors

Id	Area	Distance A	Distance B	Distance C	...	Distance M
1	mesh 1	1	0	0	...	1
2	mesh 2	1	1	1	...	2
⋮	⋮	⋮	⋮	⋮	⋮	0
n	mesh n	2	1	4	...	5

economic factors.

Table 5.3 presents an example of the matrix containing distance to socio-economic factors. In this matrix, the number highlighted in red represents that the socio-economic factor C is located between 5 and 8 km from the mesh n. In the previous example, there are n elements that represent all the combinations, and there are M socio-economic factors. In appendix A.4 an analysis between harversine, Euclidean, travel and walking distances is shown. Although in this study the harversine distance was calculated, the walking distance can be approximated by multiplying 1.2 times by the Euclidean distance [99][65][62].

The total number of unique vectors that define the residential area in the UPA's urban core of Aomori MtA is 40, this is the equivalent of 71.4% of all the vectors that define the UPA's core of Aomori MtA (Figure 5.13). According to the previously mentioned figure, 73.3% of the information corresponds to the first interval (0,1] km, 23.1% to the interval (1,3] and 3.6% to the third interval.

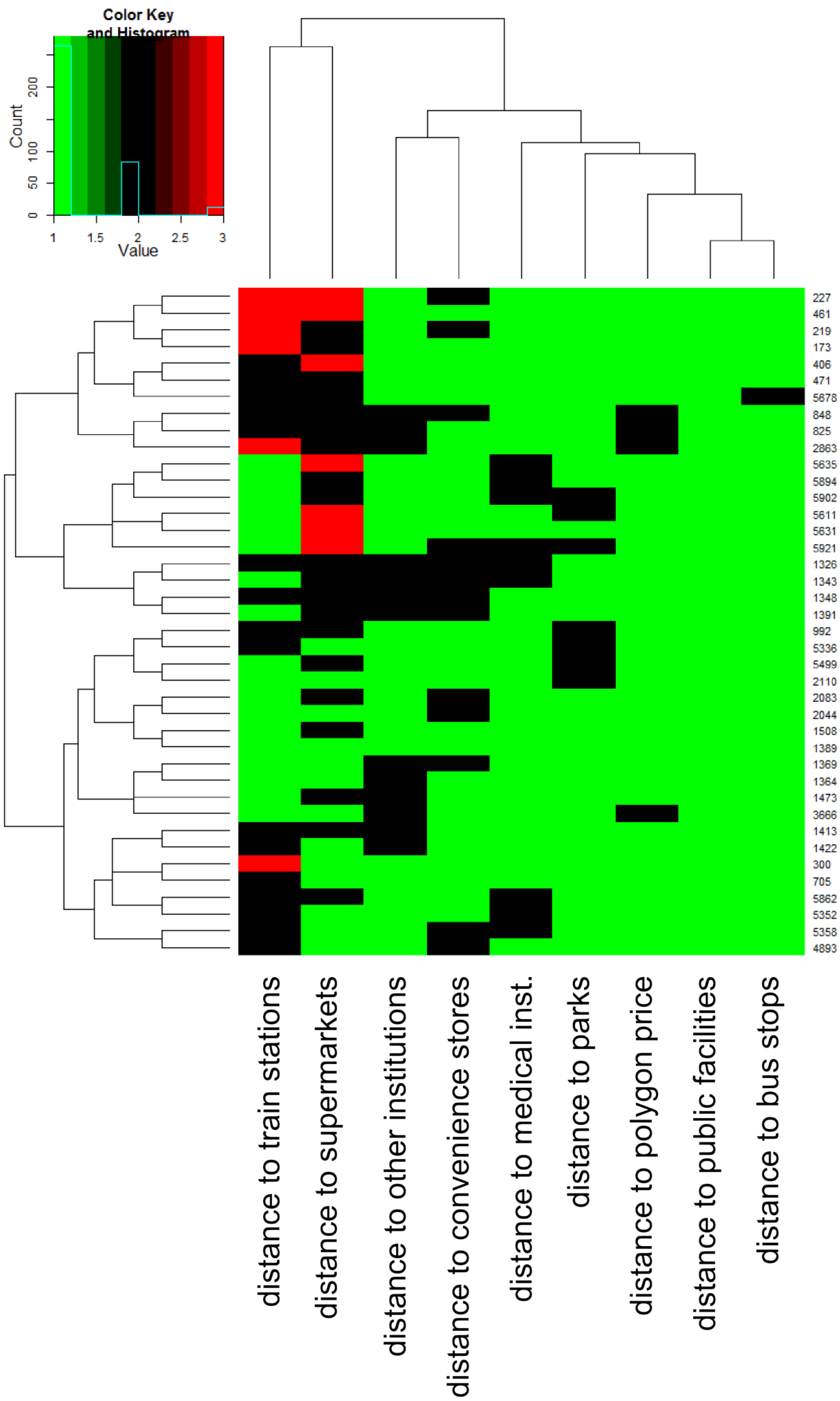


Figure 5.13: Unique vectors defining the UPA core of Aomori MtA

5.2.3 *Entropy and compactness*

The previous models developed in Chapters 3 and 4 produced the parameters necessary to create and analyze the integrated evaluation method. A summary can be seen in the Figure 5.14.

The residual Kriging model allowed to identify the urban core and suburban area of the different metropolitan areas, just the UPA of Aomori MtA has been defined properly according to the compact city model. In Chapter 3, I calculated different parameters for each MtA, parameters such as *compactness* and *entropy* are necessary to elaborate the integrated model. In Chapter 4, by using SVM, I calculated the most important related socio-economic factors for each type of land use.

From the point of view of thermodynamics, in a reversible process, the entropy can be defined using the Clausius equality. Equation (42) refers to a reversible change of a closed system, while equation (43) refers to an irreversible change of a closed system[17].

$$dS = \frac{\delta Q}{T_b} \tag{42}$$

$$dS > \frac{dQ}{T_b} \tag{43}$$

where:

S is an extensive state function, the entropy

δQ is an infinitesimal quantity of energy transferred by heat at a portion of the boundary where the thermodynamic temperature is T_b

The equation of ideal gases is defined as:

$$p = \frac{nRT}{V} \tag{44}$$

where:

n Amount of substance

R Gas constant ($8.3145\ K^{-1} mol^{-1}$)

T Thermodynamic temperature

V volume

Equation (44) shows an inverse relation between pressure and volume. There is also an inverse relation between pressure and entropy, because if the pressure is higher in a closed system the entropy and volume are reduced as well. In these experiments I could demonstrate

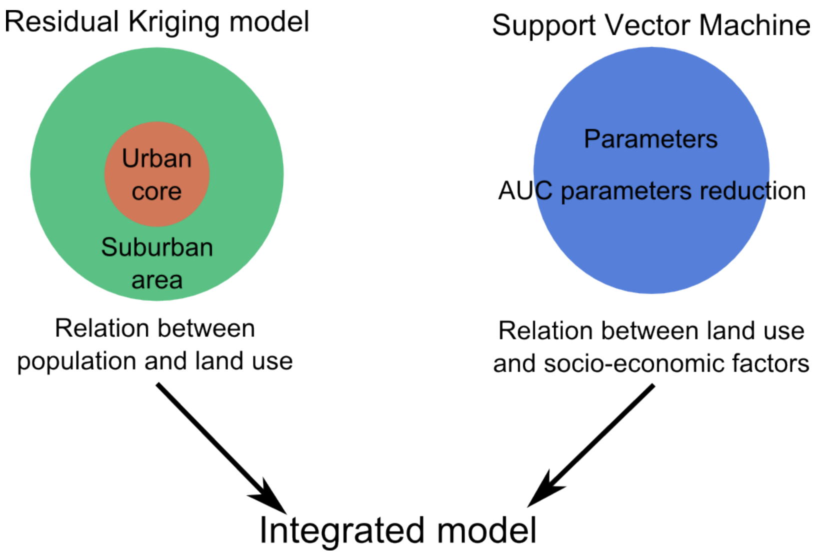


Figure 5.14: Summary of experiments

this relationship between the different metropolitan areas. In Figure 5.15, the scatterplot between compactness and entropy is shown, although the coefficient of determination is higher than 0.974 for all the models, and the slope is a negative number reflecting an inverse relation, it is suggested to gather more data in order to verify the linearity and dependency between the variables.

5.2.4 *Chemical compound*

In order to analyze the relationship between spatial points, allow me to start describing some approximations from different fields. The first one is from the point of view of mechanics explained by Newton's law of universal gravitation, defined as:

$$F = G\frac{m_1 m_2}{r^2} \tag{45}$$

Where:

F	Force between masses
G	Gravitational constant $\left(6.67384\text{x}10^{11}\text{N}\left(\frac{m}{kg}\right)^2\right)$
m_1	Mass 1
m_2	Mass 2
r	Distance between the masses

Another Approximation is given by Coulomb's law from the point of view of electromagnetism. The electrostatic force (F) acting between two electric charges can be written as:

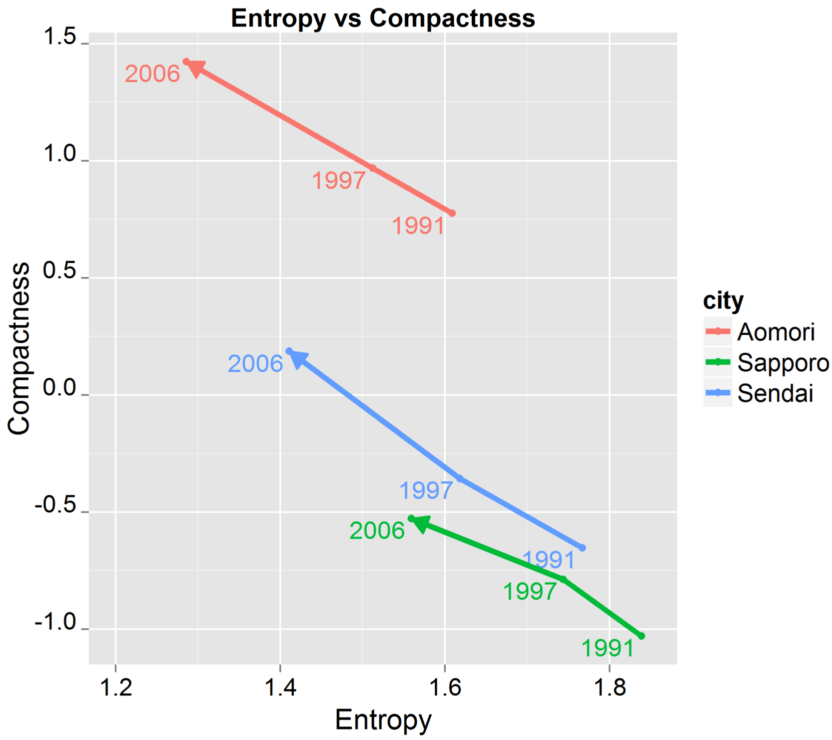


Figure 5.15: Relation between Entropy and compactness

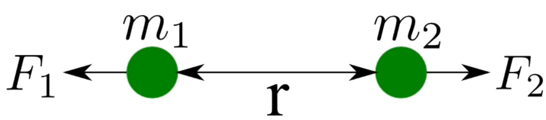


Figure 5.16: Molecular forces interaction

$$|F| = k_e \frac{|q_1 q_2|}{r^2} \tag{46}$$

$$F_1 = k_e \frac{|q_1 q_2|}{r^2} \hat{r} \tag{47}$$

Where:

F	Force between masses
k_e	Coulomb's constant $\left(8.9875x10^9 \frac{N \cdot m^2}{C^2}\right)$
q_1	Charge 1
q_2	Charge 2
r	Distance between the charges

Gravity model of trade in economics follows the equation (45) - which is similar to equation (46)- to calculate the force or interaction between cities, for instance Reilly's law of retail gravitation. This can be expressed as a ratio of the multiplied populations over the distance between each of them [35]. Other important social model involving attraction has been described by Huff [39] and it is indicated as follows:

$$P\left(C_{ij}\right) = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_j^n \frac{S_j}{T_{ij}^{\lambda}}} = \frac{\frac{A_j^{\alpha}}{D_{ij}^{\beta}}}{\sum_j^n \frac{A_j^{\alpha}}{D_{ij}^{\beta}}} \tag{48}$$

Where:

$P\left(C_{ij}\right)$	Probability of a consumer at point i traveling to retail location j
S_j	Size of retail location j
T_{ij}	Travel time from consumer at point i to j
λ	Estimated parameter to reflect the effect of travel time on different shopping trips
A_j	Measure of attractiveness of store j
D_{ij}	Distance from i to j
α	Attractiveness parameter
β	Distance decay parameter

In the equation (48), the model calculates the probability of a consumer at the point i traveling to a retail location j. This probability can be defined as $P(C_{ij}) = f(S, T)$ or $f(A, D)$.

In geography, there are different studies related to point pattern analysis; these points can be surrounded by other points representing a different set of data that may have continuous or multivariate values. A multivariate values aggregates more information than a geographic coordinates, for instance space-time, specific type of data, several variables[5]. Pairwise interaction models have been applied in different fields such as biology, chemistry and geography; these models provide patterns calculating potential functions to analyze the attractiveness and repulsiveness among all the points [57].

In this chapter a molecular structure using the model stated by Sir Lennard-Jones[41] is developed. This model calculates the interaction between two nonbonding atoms based on their distance.

$$V = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{49}$$

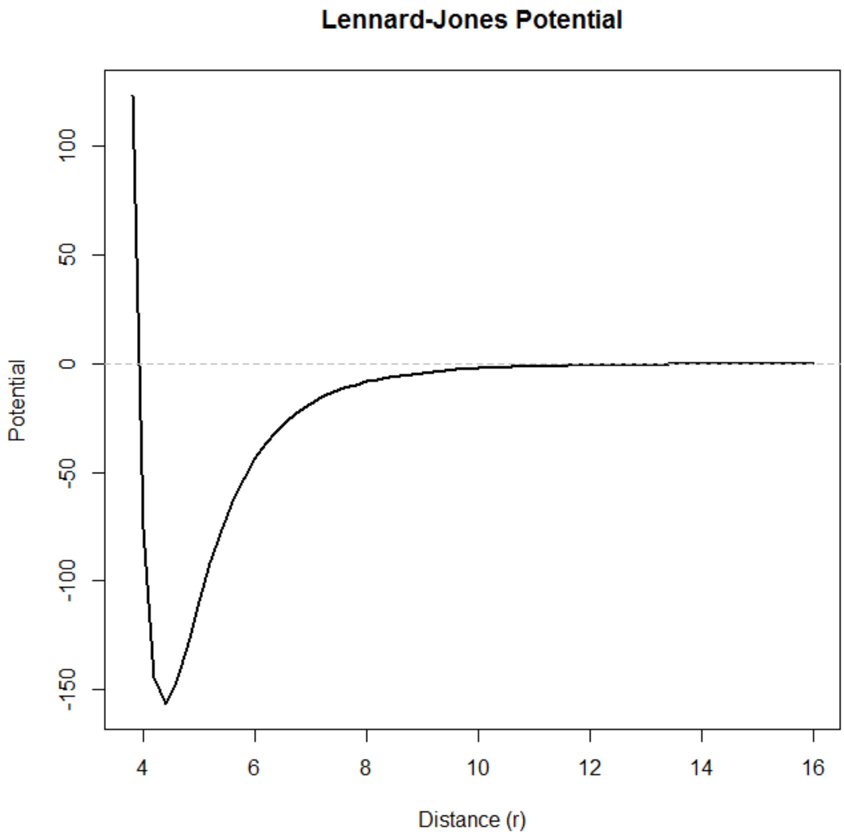


Figure 5.17: Lennard-Jones potential example

Where:

V	Intermolecular potential between two atoms
ϵ	Measure of how strongly the particles attract each other
σ	Distance at which intermolecular potential between the particles is zero
r	Distance of separation between the atoms

In equation (49) the term r^{12} refers to the repulsive term, while r^6 is the attractive long-range term.

Using the set of data that lies in the UPA's core of Aomori MtA, I proceed to evaluate the atomic interrelations between land use and socio-economic factors. According to the UPA's goal, the residential area should be promoted; for that reason, the efforts to calculate the atomic interrelations or bonds are addressed with this type of land use and the socio-economic factors.

All interrelations between the factors and land use were calculated. The total number of combinations was equal to $\binom{9}{2} = 45$.

Later, so as to calculate the distances between the different points; the coordinate system used in this study (WGS 84) using latitude and longitude was changed into Universal Transverse Mercator coordinate system also known as UTM in the zone 54. This zone corresponds to the area where the MtAs are located.

To avoid calculation errors related to redundancies, the new set of pairwise data was revised to extract possible points that may lie in the same origin.

Once the new set has been checked, the nearest neighbor was calculated to define the distance by which the window is set up for the border correction. The next argument to be defined is related to the number of iterations in order to make the algorithm converge; for this study, I have defined the maximum number of iterations same as 10,000 using a generalized linear model. Results of these experiments are shown in Tables 5.4 and 5.5, and the abbreviations used are the followings:

Factor	**Abbreviation**
Residential area (Land use)	lu
Bus stops	Bus st.
Convenience stores	Conv.st
Medical institutions	Med.ins
Other institutions	Oth.ins
Public facilities	Pub.fl
Supermarkets	Spmk
Train stations	Tr.st

In Figure 5.18 a two dimensional model is shown, however a three dimensional structure could be elaborated using all the intemolecular potential that have been calculated.

Table 5.4: Lennard Jones results experiments

Factor 1	Factor 2	Lennard Jones interaction						
		Intercept	Interact.1	Interact.2	Beta 1	Beta 2	Sigma	Epsilon
lu	Bus st.	-11.115	1.670	4.854	1.49E-05	1.49E-05	0.992	3.528
lu	Conv.st	-11.109	-0.001	-2.795	1.50E-05	1.50E-05	1.928	-2,471.126
lu	Med.ins	-11.036	1.761	4.417	1.61E-05	1.61E-05	1.360	2.770
lu	Oth.ins	-11.116	0.000	-0.383	1.49E-05	1.49E-05	1.016	-193.272
lu	Parks	-11.098	-0.001	-4.375	1.51E-05	1.51E-05	2.241	-3,404.368
lu	Price	-11.112	0.000	-0.389	1.49E-05	1.49E-05	1.016	-198.014
lu	Pub.fl	-11.035	0.000	-0.057	1.61E-05	1.61E-05	0.618	-22.316
lu	Spmk	-10.988	-0.001	-3.517	1.69E-05	1.69E-05	9.134	-3,295.643
lu	Tr.st	-11.099	-0.006	-12.609	1.51E-05	1.51E-05	3.783	-6,978.116
Bus st.	Conv.st	-13.572	-38,973.780	-38,970.159	1.28E-06	1.28E-06	4.052	-9,741.635
Bus st.	Med.ins	-12.946	285.976	290.335	2.39E-06	2.39E-06	1.581	73.690
Bus st.	Oth.ins	-13.604	5,289.480	5,292.845	1.24E-06	1.24E-06	4.052	1,324.053
Bus st.	Parks	-13.488	-38,550.872	-38,547.623	1.39E-06	1.39E-06	4.052	-9,636.093
Bus st.	Price	-13.556	5,225.377	5,228.694	1.30E-06	1.30E-06	4.052	1,308.003
Bus st.	Pub.fl	-12.945	28.044	30.851	2.39E-06	2.39E-06	3.988	8.485
Bus st.	Spmk	-13.777	-35,364.449	-35,360.911	1.04E-06	1.04E-06	4.052	-8,839.343
Bus st.	Tr.st	-13.792	-38,923.917	-38,920.364	1.02E-06	1.02E-06	4.052	-9,729.203
Conv.st	Med.ins	-13.030	2.921	7.289	2.19E-06	2.19E-06	1.361	4.547
Conv.st	Oth.ins	-14.334	6.601	9.572	5.96E-07	5.96E-07	59.892	3.470
Conv.st	Parks	-14.111	3.215	4.325	7.44E-07	7.44E-07	66.953	1.455
Conv.st	Price	-14.273	1.420	3.778	6.33E-07	6.33E-07	54.125	2.513
Conv.st	Pub.fl	-13.210	36.398	39.447	1.83E-06	1.83E-06	5.803	10.688
Conv.st	Spmk	-14.777	2.880	5.872	3.82E-07	3.82E-07	78.319	2.993
Conv.st	Tr.st	-14.828	4.384	6.714	3.63E-07	3.63E-07	82.188	2.571
Med.ins	Oth.ins	-13.099	-0.780	3.739	2.05E-06	2.05E-06	-	-
Med.ins	Parks	-13.042	323.868	328.436	2.17E-06	2.17E-06	1.582	83.267
Med.ins	Price	-13.153	303.137	307.347	1.94E-06	1.94E-06	1.582	77.904
Med.ins	Pub.fl	-12.684	10.256	14.126	3.10E-06	3.10E-06	1.341	4.864
Med.ins	Spmk	-13.115	299.821	304.032	2.01E-06	2.01E-06	1.582	77.075
Med.ins	Tr.st	-13.138	272.034	276.373	1.97E-06	1.97E-06	1.581	70.196
Oth.ins	Parks	-14.237	2.861	4.678	6.56E-07	6.56E-07	64.809	1.913
Oth.ins	Price	-15.894	0.000	0.090	1.25E-07	1.25E-07	30.159	5.472
Oth.ins	Pub.fl	-13.246	4.453	8.108	1.77E-06	1.77E-06	5.322	3.691
Oth.ins	Spmk	-14.798	2.024	5.226	3.74E-07	3.74E-07	61.523	3.373
Oth.ins	Tr.st	-14.871	1.180	2.107	3.48E-07	3.48E-07	168.635	0.940
Parks	Price	-14.158	3.855	5.915	7.10E-07	7.10E-07	61.584	2.269
Parks	Pub.fl	-13.163	11.384	14.847	1.92E-06	1.92E-06	5.627	4.841
Parks	Spmk	-14.458	2.622	4.907	5.26E-07	5.26E-07	63.366	2.296
Parks	Tr.st	-14.575	4.881	7.902	4.68E-07	4.68E-07	64.917	3.198
Price	Pub.fl	-13.216	36.407	39.440	1.82E-06	1.82E-06	5.803	10.681
Price	Spmk	-14.666	0.753	2.151	4.27E-07	4.27E-07	60.500	1.537
Price	Tr.st	-14.727	1.723	2.533	4.02E-07	4.02E-07	132.731	0.931
Pub.fl	Spmk	-13.352	38.935	41.618	1.59E-06	1.59E-06	5.816	11.122
Pub.fl	Tr.st	-13.363	38.951	41.645	1.57E-06	1.57E-06	5.816	11.131
Spmk	Tr.st	-15.990	0.911	1.881	1.14E-07	1.14E-07	291.171	0.971

Table 5.5: Strauss results experiments

Factor 1	Factor 2	Strauss interaction				Intermolecular potential
		Intercept	Interaction	Intensity	Closest distance	
lu	Bus st.	-11.114	-15.188	2.82E-05	1.18	-0.903
lu	Conv.st	-11.112	-15.190	2.82E-05	7.53	-0.001
lu	Med.ins	-11.036	-15.267	3.05E-05	1.59	-0.959
lu	Oth.ins	-11.116	-15.186	2.82E-05	3.61	-0.002
lu	Parks	-11.103	-15.199	2.84E-05	8.56	-0.001
lu	Price	-11.112	-15.190	2.83E-05	3.62	-0.002
lu	Pub.fl	-11.035	-15.267	3.05E-05	2.11	-0.003
lu	Spmk	-11.064	-16.216	2.78E-05	36.00	-0.001
lu	Tr.st	-11.118	-16.183	2.77E-05	13.66	-0.002
Bus st.	Conv.st	-13.567	-14.735	2.45E-06	4.05	0.000
Bus st.	Med.ins	-12.936	-15.366	4.61E-06	1.59	-0.059
Bus st.	Oth.ins	-13.591	-15.712	2.39E-06	4.05	-0.003
Bus st.	Parks	-13.484	-15.819	2.66E-06	4.05	0.000
Bus st.	Price	-13.543	-15.759	2.51E-06	4.05	-0.003
Bus st.	Pub.fl	-12.934	-15.369	4.60E-06	4.05	-0.331
Bus st.	Spmk	-13.770	-14.532	2.01E-06	4.05	0.000
Bus st.	Tr.st	-13.786	-14.517	1.98E-06	4.05	0.000
Conv.st	Med.ins	-13.025	-14.278	4.27E-06	1.59	-0.961
Conv.st	Oth.ins	-14.265	-15.025	1.23E-06	63.72	-0.856
Conv.st	Parks	-14.066	-15.217	1.49E-06	70.35	-0.763
Conv.st	Price	-14.237	-15.054	1.26E-06	63.71	-0.938
Conv.st	Pub.fl	-13.195	-15.107	3.59E-06	5.88	-0.285
Conv.st	Spmk	-14.680	-14.595	8.20E-07	88.19	-1.000
Conv.st	Tr.st	-14.758	-14.478	7.59E-07	88.24	-0.906
Med.ins	Oth.ins	-13.094	-15.209	3.99E-06	1.59	-
Med.ins	Parks	-13.029	-15.273	4.24E-06	1.59	-0.055
Med.ins	Price	-13.141	-15.162	3.80E-06	1.59	-0.055
Med.ins	Pub.fl	-12.679	-14.624	6.00E-06	1.41	-0.795
Med.ins	Spmk	-13.100	-15.202	3.93E-06	1.59	-0.055
Med.ins	Tr.st	-13.123	-15.180	3.84E-06	1.59	-0.062
Oth.ins	Parks	-14.179	-15.107	1.33E-06	70.35	-0.950
Oth.ins	Price	-14.759	1.781	1.34E-06	50.00	-0.183
Oth.ins	Pub.fl	-13.235	-15.068	3.45E-06	5.88	-0.990
Oth.ins	Spmk	-14.754	-14.535	7.66E-07	72.06	-0.949
Oth.ins	Tr.st	-14.777	-15.268	7.16E-07	185.74	-0.986
Parks	Price	-14.086	-15.201	1.45E-06	66.14	-0.908
Parks	Pub.fl	-13.147	-15.156	3.76E-06	5.88	-0.715
Parks	Spmk	-14.412	-14.868	1.07E-06	70.35	-0.995
Parks	Tr.st	-14.509	-14.776	9.69E-07	70.35	-0.945
Price	Pub.fl	-13.201	-15.102	3.57E-06	5.88	-0.284
Price	Spmk	-14.644	-14.635	8.49E-07	72.07	-0.910
Price	Tr.st	-14.672	-14.514	8.10E-07	141.53	-0.870
Pub.fl	Spmk	-13.335	-14.968	3.09E-06	5.88	-0.241
Pub.fl	Tr.st	-13.345	-14.957	3.06E-06	5.88	-0.242
Spmk	Tr.st	-15.879	-14.839	2.38E-07	328.55	-0.999

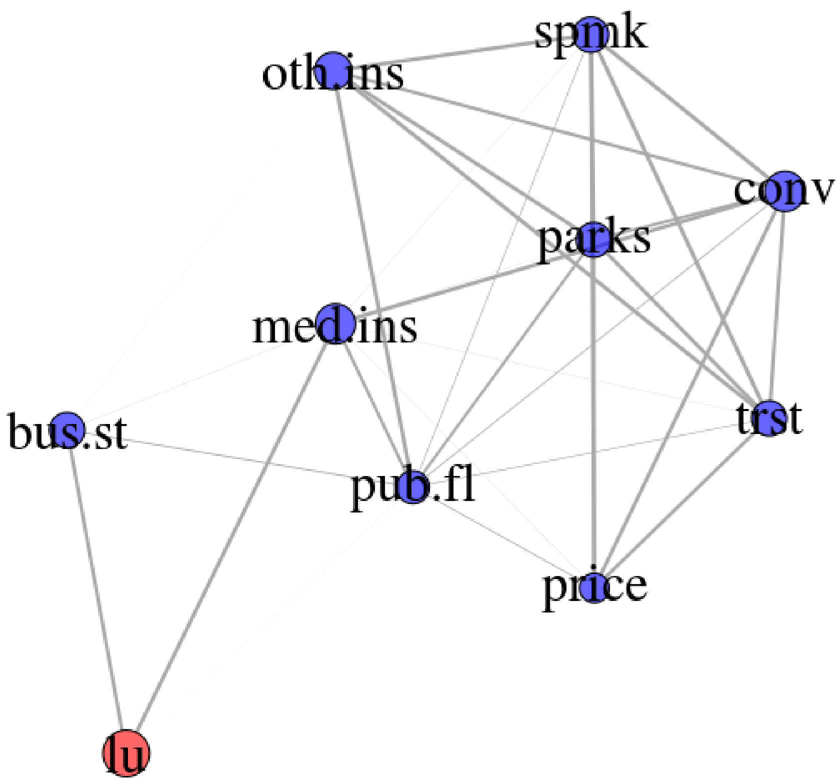


Figure 5.18: Molecule calculated for the UPA urban core of Aomori MtA using residential area

5.3 RESULTS

5.3.1 *Compact city model Analysis - Aomori MtA*

In chapter 3, intervals for population density have been defined in order to identify the core of the different UPAs. In Aomori MtA, I have outlined a polygon that defines the UPA and it covers most of this area (Figure 5.19 in green). For Sendai and Sapporo MtAs (Figures 5.20 and 5.21) two areas have been analyzed in order to study the behavior in the core, those areas are displayed in gray and purple colors and represent the population density in the UPAs of the different MtAs. They will be called as urban core type 1 and 2 respectively.

Table 5.6 shows the information related to the urban core area for all the MtAs. In Aomori MtA, the urban core area corresponds to 87.2% of the total UPA while in the other MtA's urban cores do not reach 50% of the total UPA, even when a large urban core (core type 2) has been defined for the UPA. The residential area of all the MtAs occupies between 67% - 68% of the MtA's core. Using the informa-

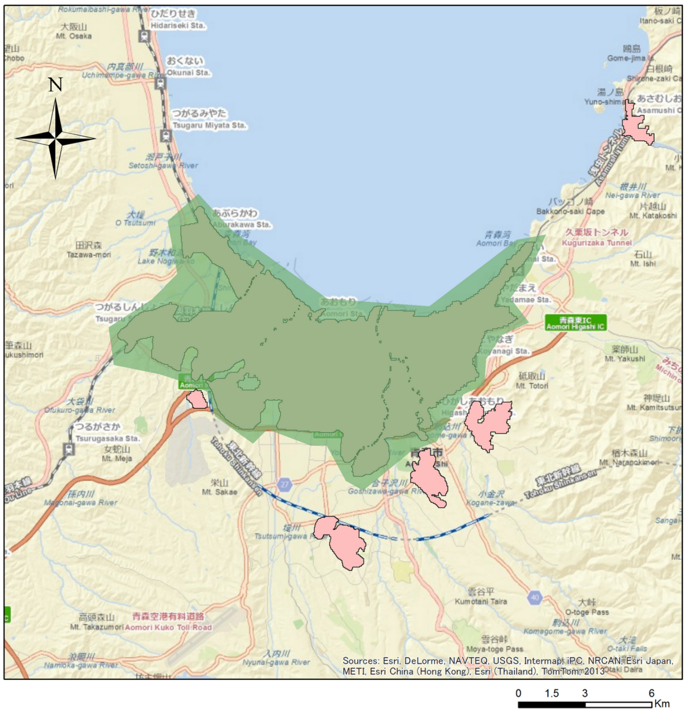


Figure 5.19: UPA of Aomori MtA for 2006 (ROI)

Table 5.6: Results in the urban core of the UPA of each MtA

MtA	Core type	Area (ha)	Percentage of total UPA	Residential area	Vectors defining R. Area
Aomori	1	5,261	87.23%	67.69%	40
Sendai	1	8,746	28.34%	67.81%	54
Sendai	2	14,905	48.30%	66.57%	67
Sapporo	1	23,625	32.48%	68.46%	72
Sapporo	2	32,695	44.94%	64.96%	100

tion from the UPA's core of Aomori MtA, it was possible to identify unique vectors for the socio-economic factors that describe the residential area. In Aomori MtA, 40 vectors were necessary to describe the residential area, while in the UPA's urban core of Sendai and Sapporo MtAs 54 and 72 vectors respectively. This value is increasing because the size of UPA's urban core and Shannon's entropy in the area are also increasing.

The coefficients of determination (r^2) between the entropy and the number of unique vectors for the urban core type 1 and 2 are 0.9995 and 0.9999 respectively. It denotes that there is a direct relation between entropy and the number of unique vectors. As well as entropy, compactness index has an inverse relation with the total number of vectors that define residential area, the coefficient of determination between compactness and urban core type 1 and 2 are 0.95 and 0.956

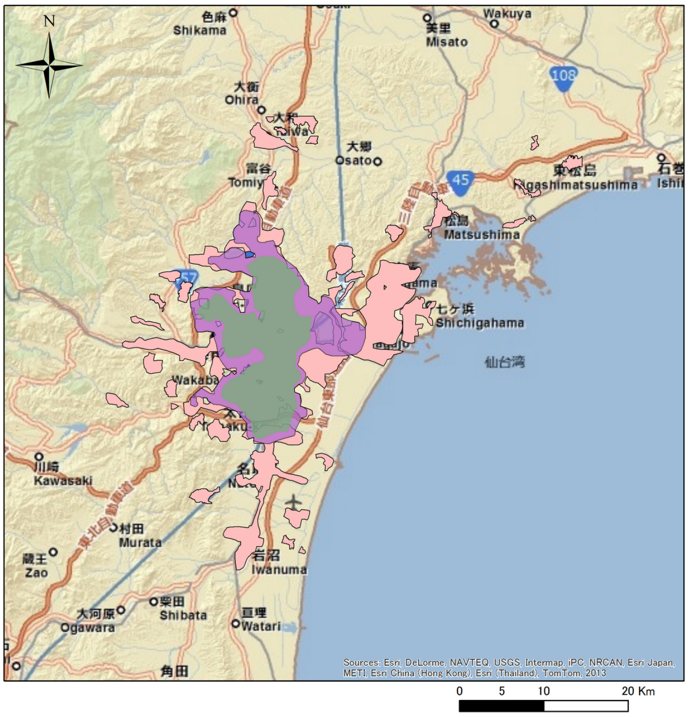


Figure 5.20: UPA of Sendai MtA for 2006 (ROI)

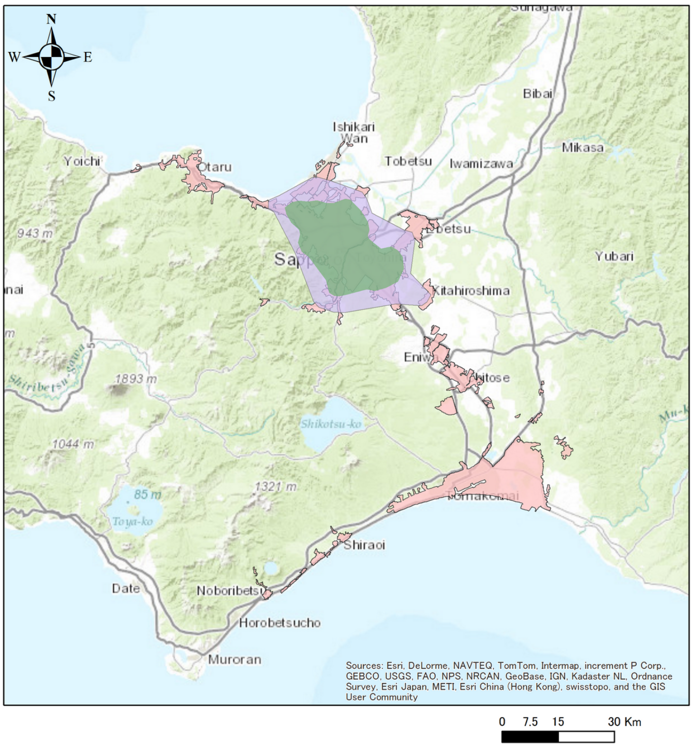


Figure 5.21: UPA of Sapporo MtA for 2006 (ROI)

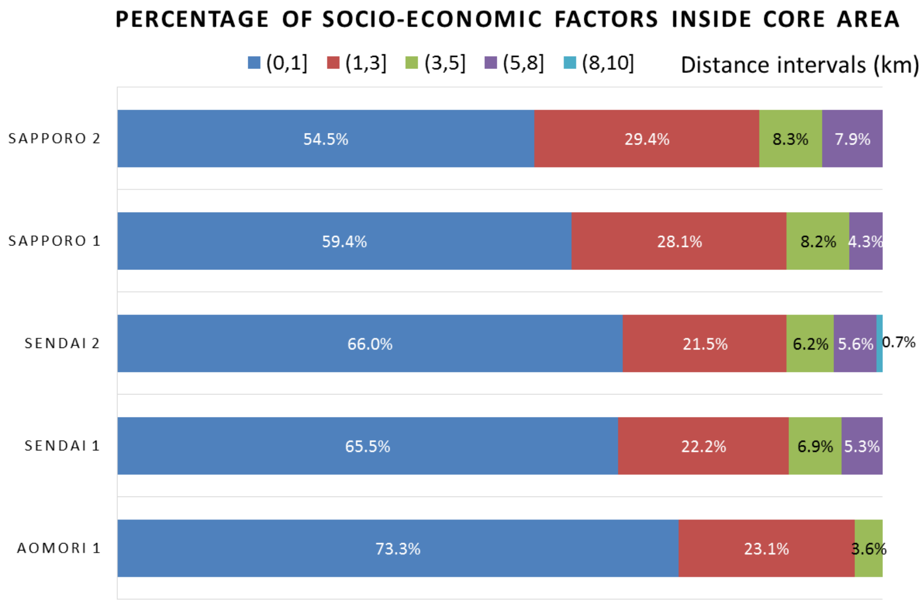


Figure 5.22: Socio economic factors located in the urban core area

respectively.

In the different MtAs it was possible to identify how many socio-economic factors are located closer to each mesh unit in the UPA. Figure 5.22 shows that in the UPA of Aomori MtA 73.0% of the socio-economic factors are nearby within 1 km, and 23.0% between 1 and 3 km. However, for the UPA of Sendai MtA the behavior between the areas defined as urban core 1 and 2 is almost same. Between 65.5% and 66.0% of the factors are between 0 and 1 km and almost 22.0% of them between 1 and 3 km. For the UPA of Sapporo MtA, 59.4% of the factors are nearby within 0 and 1 km, and there is a difference of 5.0% with urban core number 2.

In order to integrate the last chapters allow me to summarize some relations:

1. The slope between entropy and compactness indexes is less than 0. (Chapter 3).

$$\frac{\Delta Entropy}{\Delta Compactness} < 0 \tag{50}$$

Where:

Entropy	f(Land use)
Compactness	f(Land use)

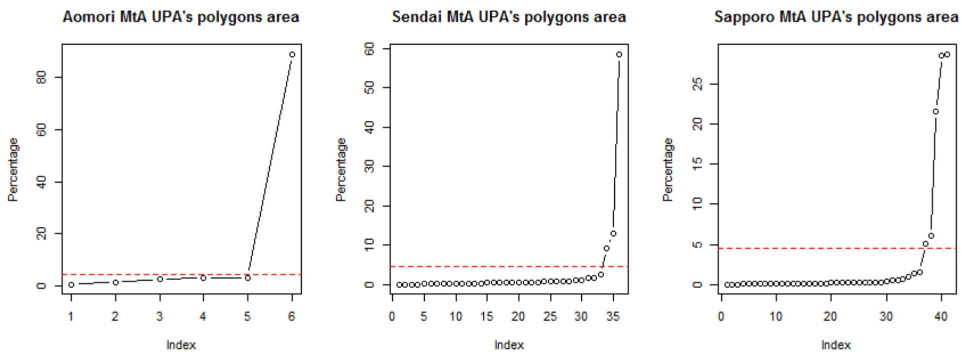


Figure 5.23: Polygons area threshold

2. Entropy is proportional to the number of unique vectors of Socio-economic factors that define the UPA's urban core of the MtA (Chapters 3, 4 and 5).

$$\text{Entropy} \propto \text{No.vectors} \tag{51}$$

Where:

No. vectors	f(Socio-economic factors, Land use)

3. The urban core area is proportional to the number of unique vectors of Socio-economic factors that define the UPA's urban core of the MtA (Chapter 3 and 5).

$$\text{CoreArea} \propto \text{No.vectors} \tag{52}$$

Where:

Core Area	f(Population density)

4. A new factor has to be defined as the proportion of the large detached areas with the UPA, it will be named as polygon compactness (PC) (Chapters 4 and 5):

$$\text{PC} = \frac{1}{\text{core.value} + \text{detached.areas}} \tag{53}$$

Where:

Core.value	Represents the number of defined urban core areas in the UPA
Detached.areas	Represents the number of polygons with more than 4.5% of the total UPA area

The previous percentage (4.5%) was extracted analyzing all the MtAs and identifying the threshold for the polygons area, it is shown in Figure 5.23.

5. There is an inverse relation between entropy and polygon compactness, PC has been defined in equation (53).

$$\mathrm{Entropy} \propto \frac{1}{PC} \tag{54}$$

Because the compactness and entropy have an inverse relation, I take into account the entropy to define a relation between population density, land use and socio-economic factors. This is to reduce the number of factors in the equation.

Integrating equations (50), (51) and (53), I obtain:

$$\mathrm{No.Vect} * A = B * \mathrm{Entropy} * PC \tag{55}$$

According to the experiments and for the sake of obtaining a relation between the factors before mentioned, I will clear "A" in the equation (55):

$$A = B * \frac{\mathrm{Entropy} * PC}{\mathrm{No.Vect}} \tag{56}$$

The term "B" on the equation (56) is a correction factor, however, in this analysis will be same as 1. The equation to calculate the behavior of the compact city model with variables related with land use, population density and socio-economic factors is described as follows.

$$\mathrm{Pr.mod} = \mathrm{Core.Area} * e^{\frac{\mathrm{Entropy} * PC}{\mathrm{No.Vect}}} \tag{57}$$

Where "Pr.mod" represents a preliminary model. Later, in order to reduce the range of the results I calculate the logarithm of the last equation.

$$\mathrm{Pr.mod} = B * \log\left[\mathrm{Core.Area} * e^{\frac{\mathrm{Entropy} * PC}{\mathrm{No.Vect}}}\right] \tag{58}$$

In equation (58), the term "B" corresponds to the percentage of the central urban core area in the UPA, it is a penalty factor in the compact city model. "Pr.mod" corresponds to the compact city model evaluation value, defined as **Acid value**.

$$\mathrm{Acid.value} = \mathrm{Perc.UPA} * \log\left[\mathrm{Core.Area} * e^{\frac{\mathrm{Entropy} * PC}{\mathrm{No.Vect}}}\right] \tag{59}$$

In Table 5.7 the results for the acid value are shown, however, in order to interpret those values it is necessary to calculate the limit of the equation for each MtA. Table 5.8 shows the best values for each MtA.

The specific characteristics of each MtA define the success of the compact city model apply to them. For that reason and considering that the total area of the urban core remains constant, the best value for each metropolitan area is shown in Table 5.9. In the previous table, it is clear that the percent completion of the compact city model applied in the UPA of Aomori MtA is 87.3%. However, in Sendai and Sapporo MtAs, the percent completion is 28.0% and 32.0% respectively.

Sendai and Sapporo MtAs need to continue working on concentrating the citizens in the urban core, preventing urban sprawling, improving the socio-economic factors in their UPA and improving the land use to reduce the entropy. Two of the most important problems that Japan faces are those related to aging and depopulation, for that reason the UPAs should be improved to provide enough support to the community and thus encourage housing and tourism in these areas.

Table 5.10 shows the different values related to compactness. Compactness presents the Burton's indexes for each MtA, by calculating the relation between population and land use area. The UPA of Aomori MtA presents the highest value (1.42), followed by the indexes of the UPAs of Sendai and Sapporo MtAs respectively. The proposed analysis method (acid value) takes into consideration also socio-economic factors among other factors. The results reflect that the UPA of Sapporo MtA shows a higher value than the UPA of Sendai MtA. This is because of the population and socio-economic factor distribution. One of the problems that Burton's index presents related to compactness is that, it is not possible to calculate the threshold of this value, however it is possible to compare this value in different periods of time in order to understand the progress of compactness in the urban area.

Table 5.7: Acid value calculated for the UPA of the MtAs

MtA	Acid value
Aomori	7.482197
Sendai	2.543886
Sapporo	3.223805

Table 5.8: Best parameter values for indices

Factor	value	Description
Perc.UPA	1	The urban core of the MtA is the same as the UPA
Entropy	0.08	Although Shannon's entropy threshold is defined as 1.0414, however the entropy tends to 0. The value 0.08 comes from the minimum entropy value calculated by Ramachandra. This value could be chosen by the researcher or city planner.
PC	1	The urban core is same as the UPA and hence detached large urban areas are equal to 0.
No.Vect	40	The number of combinations explains the residential area for the compact city model in the UPA of Aomori MtA.

Table 5.9: Acid value calculated with best values for the UPA of the MtAs

MtA	Acid value
Aomori	8.570076
Sendai	9.078352
Sapporo	10.07206

5.3.2 *Generalization*

Because Japan has strict regulations in terms of the land use, and it also develops large quantity of data, it is difficult for other countries to apply the evaluate method which I propose in this study to their metropolitan areas. However, it is possible to use Densely Inhabited Districts (DID) in order to identify the core of metropolitan areas and detached urban areas in other countries. It is necessary to collect the information related to the land use in order to calculate the Shannon's entropy, and socio-economic information to calculate the number of vectors to describe the residential area.

Table 5.10: Comparison between Burton's compactness index and acid value

UPA	Compactness	Acid value
Aomori	1.42	7.48
Sendai	0.19	2.54
Sapporo	-0.53	3.22

The generalized model to evaluate a compact city model can be written as:

$$\mathrm{Acid.value} = \sum_{i=1}^{n} \left[\mathrm{Perc.UPA} * \log \left[\mathrm{Core.Area} * e^{\frac{\mathrm{Entropy} * \mathrm{PC}}{\mathrm{No.Vect}}} \right] \right]_i \quad (60)$$

where:

n is the total number of cores in the MtA

There is a limitation in the data acquisition, because there are available satellite data and printed maps in other countries; however, there is not enough detailed information related to socio-economic factors. In order to evaluate compact city model, it is required to collect information about the parameters involved in this study to promote the residential area in the urban area.

Adding a summation in Equation (59), it is possible to evaluate different scales of MtAs; however, it is necessary to evaluate separately the urban cores (central and regional ones). For instance, Sapporo MtA has 3 urban cores respectively in Sapporo, Otaru and Tomakomai cities; so as to evaluate properly the compact city model in this MtA, it is necessary to calculate the acid value for the three areas which have each urban core separately. Once the values in each of the three areas are calculated, the new acid value for Sapporo MtA will rise. Finally, based on the equation (60), evaluating the three areas respectively, it will be possible to decrease errors.

5.3.3 *Examining the validity of the proposal of an evaluation method for a compact city model*

In order to validate if the evaluation method proposed for a compact city model was correct, there are 2 ways: The first one is calculating the best parameters for the acid calculation in order to estimate the percent completion of the compact city model; also by comparing the compactness index and the acid value.

The second way is to conduct a technical hearing from experts. So as to confirm the validity of this proposal, the second way was developed as well. The experts' validation and comments about the proposal presented in this document are shown in Table 5.11.

A total of 12 experts on different areas such as architecture, urban planning, civil engineering, environmental sciences, geography, among others were consulted. All of them were agreed with this proposal, however, they consider that the some points should be clarified in order to develop an accurate proposal. The most relevant comments are explain from subsection 5.3.3.1 to subsection 5.3.3.8.

Table 5.11: Validation from technical hearing

ID	Professor	Specialty	Comments	Approves or rejects
1	Ikujiro Wakai	Environmental Science	Shannon's entropy and Burton's compactness are correlated.	Approves
			The importance of traveling and walking distances in a compact city model are appropriate for this evaluation method proposal.	
			In the future, it is required to deep into energy transmission and triple bottom line as sustainability of business decisions.	
2	Yuji Murayama	Geography	He pointed out that the r-squared of the linear model regression between Shannon's entropy and Burton's compactness is close to 1, however there are just 3 points for this evaluation method	Approves
3	Jun Izumi	Urban Planning	Compactness index has two aspects: Social compactness can be evaluated using daily input-output information, and systematic compactness. For that reason, as a future work he considers that I should deep into the importance of the energy efficiency and commuter traffic.	Approves
4	Hirohito Kuse	Civil Engineering	He mentions that the model could change by using different data, such as population density. He considers the importance to calculate the personal trip to the different facilities; also the importance of the energy reduction.	Approves
			He considers the importance of the Central Business District as an area that affects commercial and social activities. Also the importance to study the building capacity by using three dimensional information.	
			He stresses on the hospital allocation in a metropolitan area with a compact city model.	
5	Hideki kaji	Urban Planning	He stresses on the importance between Burton's compactness and Shannon's entropy as appropriate indexes to evaluate population distribution and urban sprawling.	Approves

Continued on next page

Table 5.11 – *Continued from previous page*

ID	Professor	Specialty	Comments	Approves or rejects
			He mentions the importance of the relation between compactness and quality of life in a compact city model.	
6	Yoshitsugu Aoki	Architecture	Makes emphasis on the data acquisition. He wants to know the importance and meaning of Shannon's entropy for this evaluation method.	Approves
			He mentions how the method could change if the researcher uses a different scale of mesh area, for example 200m mesh. He pointed out how the support vector machine model also can change the results if the inputs are different.	
			He pointed out that if data change, the model changes as well.	
7	Toshiaki Ichinose	Environmental Science	He pointed out the importance of minimizing energy use, especially in traffic fulfilling the objective of low carbon city.	Approves
			He mentions why is compactness so important for the evaluation of a compact city model.	
8	Changi Kim	Urban Planning	He considers that all the indexes should not be together in the same equation.	Approves
			It is necessary to study CO_2 emissions, calculating its index, and extract correlations with other indexes.	
9	Seiichi Kagaya	Civil Engineering	He considers that the city charm degree is required, not only population and land use.	Approves
			Why I compare compactness and entropy?.	
			How to decide the intervals of distance?.	
10	Hideharu Morishita	Environmental Science	How did I choose the distance interval?.	Approves
			Why did I evaluate the core area by ha?.	
			How can I explain the urban core and compactness value when there are multicores? .	
11	Yoji Kwakami	Civil Engineering	No comments	Approves
12	Hiromichi Fukui	Environmental Science	No comments	Approves

5.3.3.1 *Why is compactness important?*

The term compactness has been studied by different authors, however, there are different approaches involving housing density, mix of residential and nonresidential uses, street network patterns, housing structure type mixture [101] among other variables. Compactness allows effective use of green space and farmland protection, an appropriate use of land and infrastructure and a reasonable allocation of public investment [12]. Burton suggests that the compact city should encourage a fair distribution of costs and benefits if the "urban compactness is associated with benefits for the conditions of life chances of the disadvantaged", thus reducing the difference in social classes [10]. Urban compactness also influences the social equity, the previous author found that 9 social equity effects are related to urban compactness; those are: access to superstores, green spaces, public transport use, extent of walking and cycling, domestic living space, death rate by: mental illness and respiratory diseases, crime and social segregation. The before mentioned factors were studied in the U.K., however, it is possible to see that in Japan, and according to the present study and related bibliography, access to supermarkets, access to green spaces, public transportation (bus, trains) and health facilities are common factors to these countries.

5.3.3.2 *Is there a relationship between entropy and compactness?*

There are different approaches to the compact city model from the point of view of compactness. The urban compactness takes into account three aspects such as density, mix of uses and intensification. This index measures the concentration in the urban area, however Shannon's entropy is an index useful to describe the sprawl. In the approach given by Tsai et al., they compare different indexes to quantify the urban form and demonstrate that Shannon's entropy is superior to measuring sprawl than other indexes such as Gini and Delta among others. This situation is because the index is not affected by size, shape and number of sub-areas [85]. Since the relation between compactness and Shannon's entropy was not elaborated, I calculate this relationship. From the point of thermodynamics, the relation between entropy and pressure is inverse in a closed system; since the UPA did not change in the different periods of time, the hypothesis was that the relation between this two indexes could be described.

5.3.3.3 *Compactness and Quality of life*

In 1970s, the compact city concept was advocated. Some scientists said that extreme concentration makes it possible to achieve sustainability and contributes to the preservation of the environment, however, others were skeptical because there were not enough freedom of human nature, quality of life and lifestyle[33].

The urban center basically offers offices, housing and commercial functions; less diversity in residents, local services and jobs may cause an inadequate use of urban land and infrastructure. Quality of life is important for the attractiveness of the urban core. In the compact city model it is important the adequate use of public parks and green spaces. However, it is possible to promote the mixed land use in order to attract residents and socio-economic factors to the urban center, and thus promote walking and cycling environments. Some authors are concerned with traffic congestions, housing affordability, loss of open and recreational spaces, energy optimization and greening of built-up areas[1].

The quality of life has been defined in different ways, however, they have in common 3 main aspects such as objective life conditions, subjective feeling of wellbeing and personal values and aspirations. The model proposed by Felce and Perry is based on 15 sources (Figure 5.24), they define the quality of life as: "*the overall general wellbeing that includes objective descriptors and subjective evaluations of physical, material, social and emotional wellbeing together with the extent of personal development and purposeful activity, all weighted by a personal set of values*"[21].

5.3.3.4 *Why should I use 100m mesh?*

For this study the most detailed information was used. The land use was measured by the government in 100m mesh; however, the information related to population was obtained by districts. For that reason, it was necessary to construct an unified system in the same scale; in this way I applied a residual kriging model to interpolate the information into 100m mesh. The information related to land use and population correspond to geographic vector defined as polygons (area), and the information of socio-economic factors are points.

5.3.3.5 *Traffic network and travel distances*

The traffic network is relevant in the compact city model. This is because it should be optimized in order to reduce the gas emissions and energy in order to reach a destination. Yamamoto et al. [99], Newell [62], among other authors have demonstrated that it is possible to estimate a travel distance by using the Euclidean distance. The euclidean and haversine distances tend to be similar when the distances are not so long, because of the calculation method. This relation can be written as: walking distance could be expressed as 1.2 times the haversine distance. The experiments have demonstrated that the value is between 1.204 and 1.250; in our experiments I calculated this value in the UPA of Aomori's MtA same as 1.271.

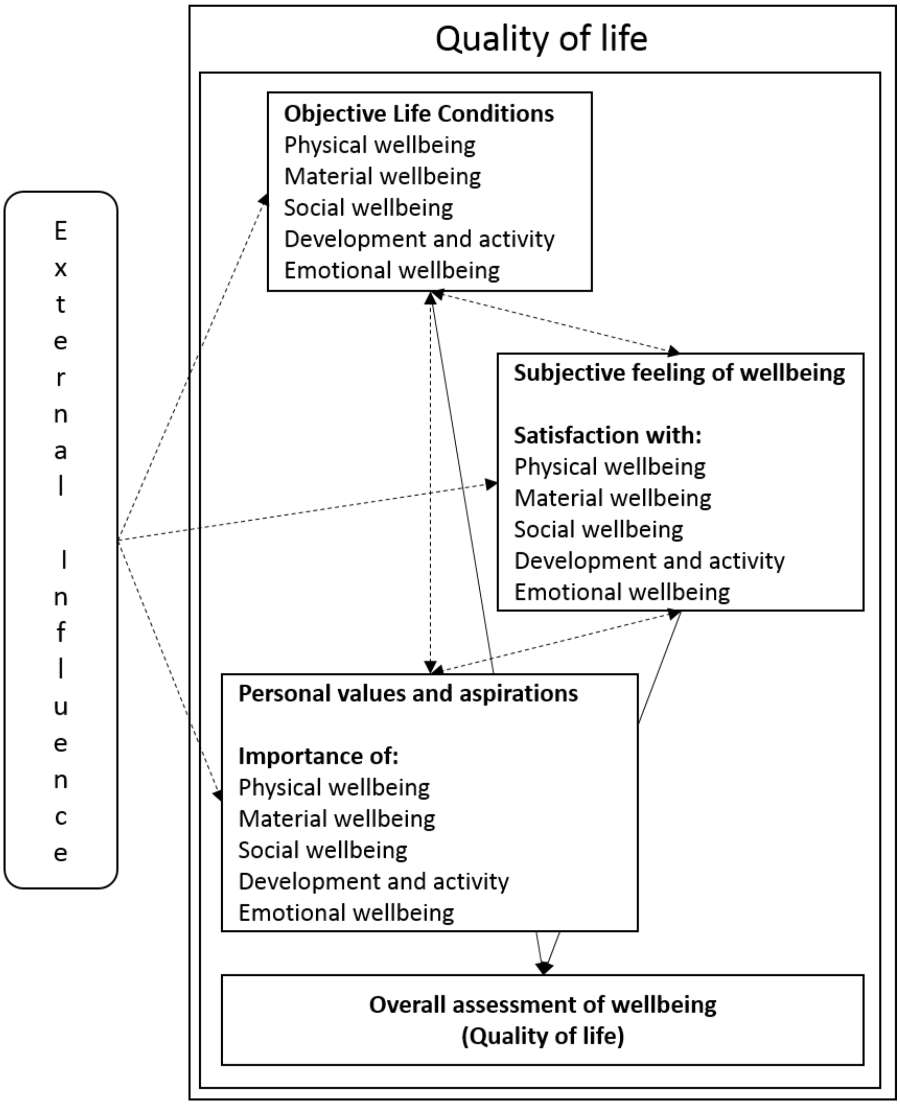


Figure 5.24: Model of quality of life (Felce et al.)

5.3.3.6 *Study of energy*

One of the most recurrent topics related to compact city model is energy. Currently the MLIT promotes eco-town and makes emphasis in low carbon emissions. In the document "Is your city an eco-town? - Let's start developing a compact city -" written by the MLIT, they are promoting the use of public transportation because the citizens can reach in this way the urban and regional cores of the UPA.

The low carbon city act (Eco-city act) was established in order to promote a friendly environment for families and citizens. Figure 5.25 shows the basic policies for the promotion of a low carbon city. For residents there are some economical incentives such as a reduction of income tax if the residence fulfills the standards of low-carbon residence; the other aspect is a special bonus to total floor-area ratio. Two

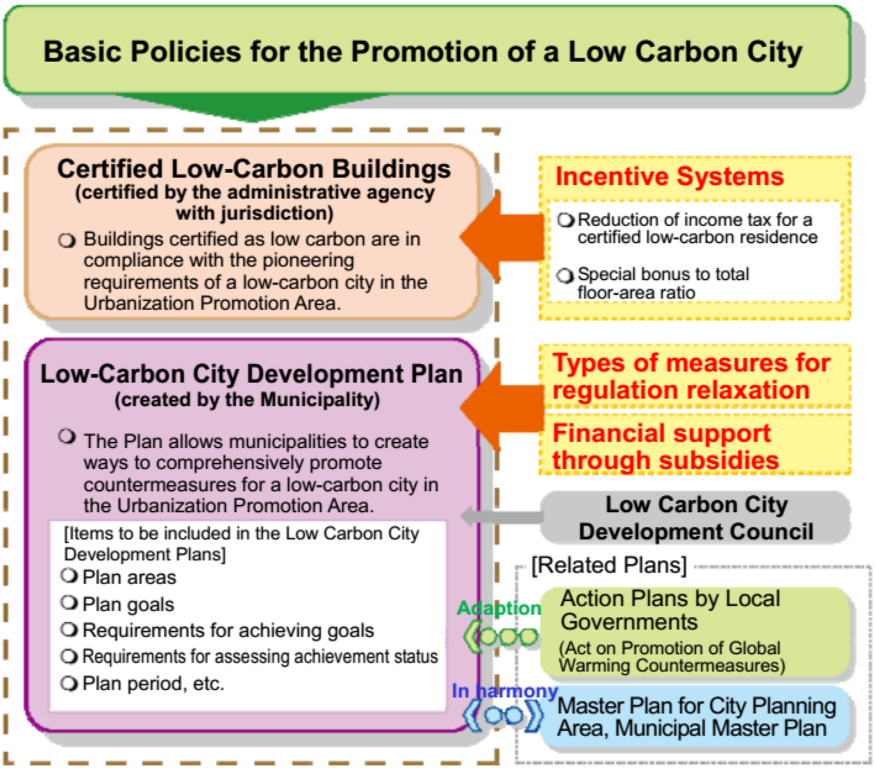


Figure 5.25: Basic policies for the promotion of a low carbon city (Taken from "Is your city an eco-town?" by MLIT)

other important aspects are the certification of low-carbon buildings for those buildings that satisfy the requirements of the low-carbon city in the UPA. Also the plan related to reach the goal in the development of low-carbon city; this plan is developed by the municipalities in order to promote the residential area fulfilling the new standards in the UPA.

5.3.3.7 *Buildings (Three dimensional data)*

The population distribution is the first aspect taken into account in the study of the compact city model. However, there are two different types of population in a region of interest such as daytime and nighttime populations. At daytime some citizens travel from their homes to their jobs and at nighttime they return to their homes. For that reason the concentration of population changes depending on the time. In a local area, these changes are not so relevant because the distances from dwellings to jobs or schools are not so longer than in large metropolitan areas. In this way, if I take into account the capacity of buildings, I can say that the population density will not change drastically because of the characteristics of these local metropolitan areas.

5.3.3.8 *Hospitals Importance*

According to the Aomori MtA's compact city model, the hospitals are distributed in the UPA. Because the problems related to depopulation and aging it is required to improve the living environment for elderly people, this will bring independence and families can be peaceful knowing that elderly residents can attend this type of facilities in an eventual case. Nowadays, hospitals and medical institutions are important socio-economic factors, not only in Japan, but also in the U.K.[10]. For that reason, it is necessary to keep studying the best allocation for this facilities, satisfying the demand of population.

6 CONCLUSION

6.1 Residual Krigging Application to Population Density

The use of GIS in the development of this study was useful to quantify changes in land use, population data and geographic information management. It was possible to graphically understand the situation of different MtAs and to use different tools for geostatistical analysis.

The major findings of this study are summarized as follows: Using this analysis, it was possible to estimate Aomori MtA's compact city model. In future analysis of population behavior, I would like to consider socio-economic factors and other variables which may affect the decision-making process for housing.

The variogram model is useful in finding the best Kriging parameters. The size of the parameters of Sapporo MtA is clearly larger than those of other MtAs. This is because of a large total area and the geographic distribution of the population. In experiments, residual Kriging models showed a disadvantage in interpolating data close the boundaries of the map range. However, this can be solved by extending the analysis area.

The purpose of this study was to describe how the compactness indexes of three Japanese cities (Aomori, Sendai and Sapporo MtAs) have changed over time to see if they follow the objectives of a compact city. Using variograms and later the residual Kriging model, I was able to estimate the population density by 100m mesh; and in this way understand how the population is distributed in the compact city model. Through residual Kriging model calculation, it was possible to identify commuter belts in the Sendai and Sapporo MtAs.

Further improvements in the residual Kriging model are required to improve the model scope, because the model presents problems near the boundaries and for that reason it was necessary to increase the extent. In order to evaluate the compactness in a city, there are several metrics and therefore it is necessary to choose appropriately which of these are useful to each study. Improved residual Kriging models should be applied to MtAs in other countries.

6.2 SVM Application to Socio-Economic Factors

In this study I have contributed to integrate geospatial information with socio-economic factors in order to classify land use. The SVM algorithm is useful to classify land use using information of socio-economic factors in a compact city model. By tuning the model, I could find the best cost and gamma parameters. It is important to choose the best variables which affect the housing in order to understand clearly the internal situation of the city. The UPA serves its purpose by gathering most of the residents in this area. It is possible to see the implementation of the compact city model in the UPA of Aomori MtA through the land use master plan and UPA's boundaries definition.

Although in this study grid search was applied, it is required to research in other search algorithms looking for a global minimum; for instance the application of deterministic, stochastic or heuristic methods. Parallel computing has been useful to address the problem related to the global minimum. By calculating different SVM models in the different MtAs, it was possible to reduce the computational time. Improving this techniques for finding a global minimum and revising mathematical formulation, it could be possible to reduce the computational resources.

The use of the AUC helped to understand how each classifier affects the model. Using this criterion, it is possible to understand hidden characteristics about the housing decision-making process. It was clear the influence of commuter belt through the AUC calculation. The most important variables for residents are related to transportation, health, land price and services.

6.3 INTEGRATED MODEL

The study between compactness and entropy in land use was addressed to demonstrate from a thermodynamic point of view that there is an inverse relationship. This was a required step in order to introduce the city planning data into thermodynamics and subsequently in the chemical field.

In order to calculate the atomic interrelations between atoms (socioeconomic factors and land use) Newtonian, gravitational and economical models were studied. Although there are some applications from the point of view of chemistry to geographical points such as environmental and those relating to mining. However, this study is a pioneer in the study of the relationship between population density, land use and socioeconomic factors. I was able to apply chemical theory such as Lennard-Jones potential for calculating a molecular structure; however, further studies are required to construct the three dimensional structure which defines a residential area in an UPA. This would be a new way to understand the different relations in the city planning. In order to construct the three dimensional space, it may be calculated using different models to construct polyhedra, or heuristic techniques such as tabu search, swarm intelligence, neural networks among other techniques. Once the procedure to construct a three dimensional structure is elaborated, it is possible to recreate other structures for different type of land use.

Through the vectors that define the residential area was possible to create a relationship between them and other factors such as the area of the MtA core and entropy. These vectors allowed us to grasp the behavior of the socio-economic factos in the UPA. By calculating the area of the polygons that make the UPA, I could identify the threshold for a normal polygon area (4.5% of the total UPA), this value was useful to identify large detached polygons. In this study, the UPA of Aomori MtA's compact city model was analyzed, for that reason I used its present indexes to compare the model in the other MtAs.

By calculating the "acid value" in chapter 5, it is possible to integrate different aspects that had not been integrated yet, it is possible to calculate the future goal of the compact city model in an UPA; also understand how far a city or MtA is from this goal.

6.4 FUTURE WORK

In future studies other aspects should be analyzed in order to improve the accuracy for the compact city model evaluation; for instance energy efficiency, spatial distribution of urban buildings, traffic congestion among other variables. Taking into account this new variables, the evaluation model will help to improve the quality conditions, energy transmission, CO_2 reduction and security. As well as before mentioned aspects, logistics is an important aspect. Since the MtA's size increases, it is more difficult to calculate best routes, for that reason an appropriate algorithm is required. I would like to continue working on heuristic techniques in order to find faster and appropriated results.

In future studies, it is suggested to construct three dimensional molecules for all the land use types in the core area, comparing these structures with other MtAs and study how these are interacting each other.

Finally, the execution of a compact city model should be tested in other countries so as to improve the quality of life and environment.

Part I

APPENDIX

A

APPENDIX

In this part, information related with coding and GIS tools will be added. Section A.2 presents an introduction and results for the use of parallel computing by calculating SVM models. Section A.3 presents an experiment related with noise in the different MtA UPA, and it will explain why the noise as socio-economic factor was not taken into account. The section A.6 will explain how to manage the geographical data in the different parts of this study, presenting the programming code written in different software.

A.1 DETAILED INFORMATION OF SOCIO-ECONOMIC FACTORS IN AOMORI UPA

Bus stops data
Total records: 433

Train stations
Total records: 12

Convenience stores
Total records: 83
Detailed information related to convenience stores is shown in table A.1. The information was downloaded in the following link: `http://post.kiramori.net/post/cvs_city.php?pref=%E5%AE%AE%E5%9F%8E%E7%9C%8C` on 2012-07-26.

Medical institutions
Total records: 382
Detailed information related to medical institutions is shown in table A.2. The information related to medical specialties can be seen in figure A.1. According to the previously mentioned figure, the most

Table A.1: Detailed convenience stores

Shop name	Total
Circle K Sunkus	20
Sunkus	16
Ministop	7
Lawson	40

Table A.2: Detailed medical institutions

	Country	Pub.Fl	Med.Corp	Privates	Others	NA
Hospitals	1	4	6	1	8	0
Clinics	0	0	0	0	0	226
Dental clinics	0	0	0	0	0	136

Table A.3: Detailed other institutions

Type	Frequency
Farmland	1
Bank	2
Factory	2
Others	3
Office	5
Warehouse	6
Store	14
Residence	55

important specialties in the UPA are internal medicine, dentistry, pediatrics, Gastroenterology and cardiovascular areas.

other institutions
Total records: 72
Detailed information related to other institutions is shown in table A.3.

Parks
Total records: 120
Detailed information related to parks code is shown in table A.4. In table A.5 information related to parks and its planning is introduced.

Price
Total records: 89
Information related to land use price depending on social activity at point location is shown in table A.6.

Public facility
Total records: 504
Information related to the major divisions in public facilities is shown in table A.7. The administration code is also shown in table A.8. A cross tabulation between major divisions and administration code is

Table A.4: Code for parks

Code	Name
1	Block Park
2	Neighborhood Park
3	District Park (Country Park)
4	General Park
5	Athletic Park
6	Broader-based Park
7	Recreation City
8	Government-managed Park
9	Special Park (Scenic Park, Animals-and-Plants Park, History Park)
10	Buffer Green Zone
11	City Green Tract of Land
12	Green road
13	Urban Forest
14	Open Space Park

Table A.5: Parks planning

	1	2	3	4	5	6	9	11	12	14
Determined	56	9	1	2	0	2	3	13	2	1
Undecided	25	1	0	0	1	0	0	4	0	0

introduced in table A.9.

Supermarkets
Total records: 12
Information related to department stores and supermarkets by sales amount is shown in table A.10.

Table A.6: Social activity at point location

Type	Freq
Farmland	1
Factory	2
Hotel	2
Others	3
Warehouse	5
Office	6
Store	13
Residence	71

Table A.7: Public facility (major divisions)

Code	Type
3	Building
9	Other
11	National institutions
12	Local government
13	Welfare agency
14	Police agency
15	Fire station
16	School
17	Hospital
18	Post office
19	Welfare facility

Table A.8: Administration code

code	type
0	Others
1	Country
2	all prefectures
3	Cities, Wards, Towns, and Villages
4	Private sector

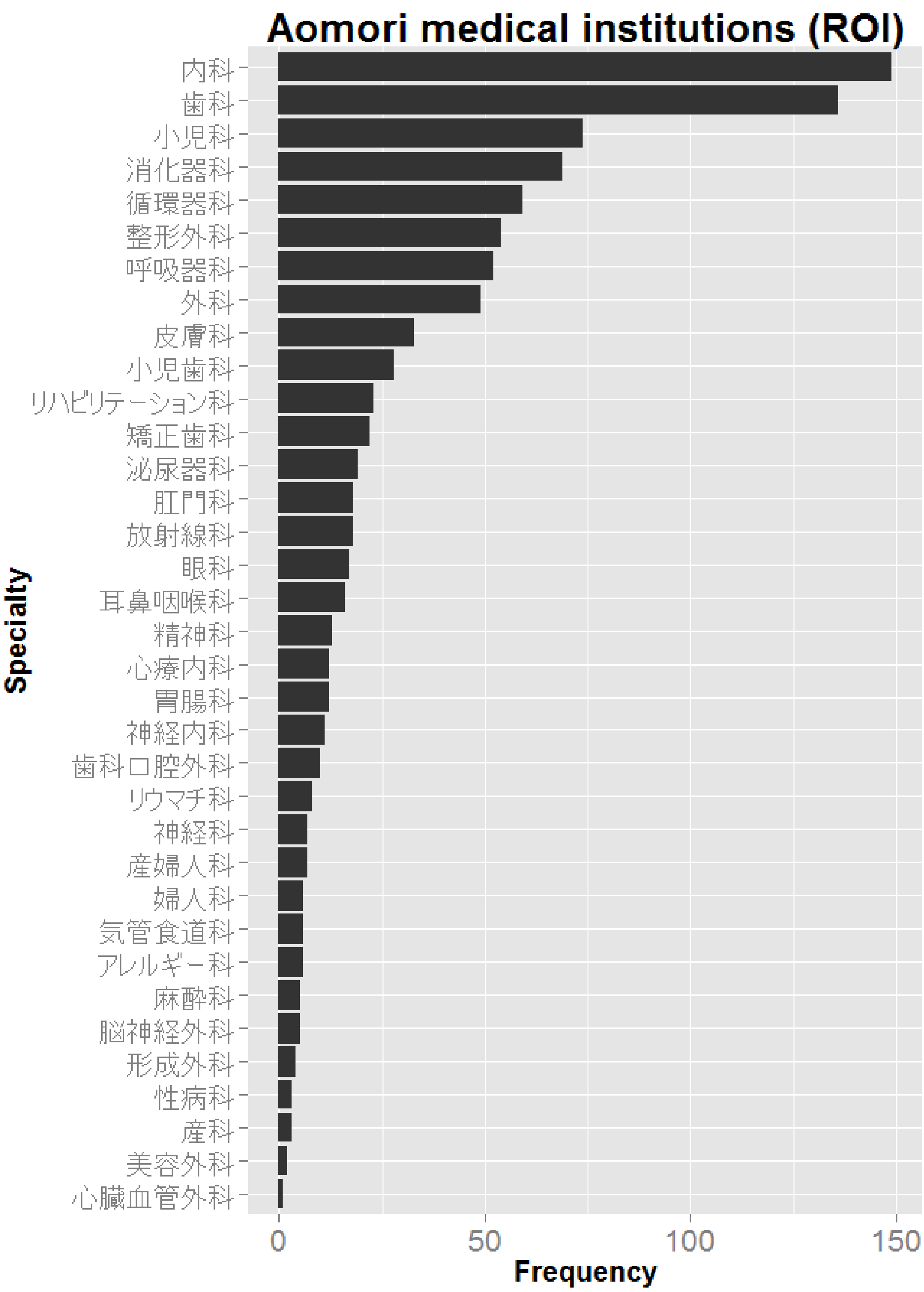


Figure A.1: Medical specialties

Table A.9: public facilities by administration codes

	3	9	11	12	13	14	15	16	17	18	19
0	0	17	0	0	0	0	0	0	0	47	4
1	0	0	36	0	0	0	0	0	1	0	0
2	5	0	0	5	1	24	0	18	3	0	8
3	4	0	0	1	0	0	11	63	1	0	70
4	2	0	0	0	0	0	0	45	15	0	123

Table A.10: number of department stores and supermarkets by sales amount

Sells	Total department stores and supermarkets
30 million - 1 billion	7
1 billion -50 billion	4
more than 50 billion	1

A.2 PARALLEL COMPUTING

Most of the software environments have been written for a serial computation[1]. This means that:

1. Are made to be run on a single computer having a single Central Process Unit (CPU).
2. The problem is broken into a discrete series of instructions.
3. The instructions are executed first one first served.
4. Only one instruction may execute at any moment in time.

On the other hand the parallel computing has the following characteristics:

1. To be run using multiple processors
2. A problem is broken into discrete parts that can be solved concurrently
3. Each part is further broken down to a series of instructions
4. Instructions from each part execute simultaneously on different processors
5. An overall control/coordination mechanism is employed

Figure A.2 and A.3 show examples of serial and parallel process. Task partitioning is representing a buying process and a cashier, this process is also called First-come first-served (FCFS) or First-in first-out (FIFO).

So as to construct a parallel environment the computer resources might be:

1. A single computer with multiple processors
2. An arbitrary number of computers connected by a network
3. A combination of both

A.2.1 *Results*

For this analysis a single computer with multiple processors was used, the CPU characteristics are shown in Table A.11.

Because of the size of the Aomori MtA UPA it was not necessary to calculate a model using parallel computing. The result comparing

1 Introduction to Parallel Computing: https://computing.llnl.gov/tutorials/parallel_comp/

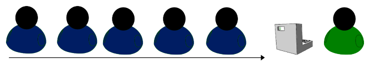


Figure A.2: Serial process example

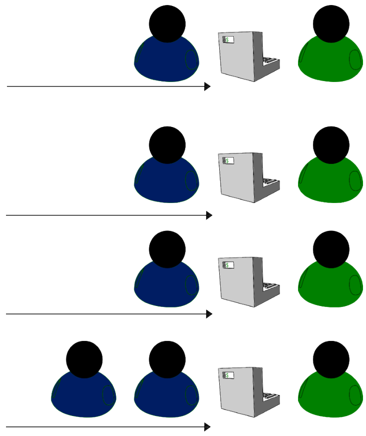


Figure A.3: Parallel process example

Table A.11: Information about computer

Type	Specifications
Operating system	Ubuntu 12.04 LTS
Memory	11.7 Gb
Processor	Intel Xeon (R) CPU E5640 @ 2.67Ghz x 4
Os type	64-bit
Disk	971.8 Gb

parallel vs. serial computing using 3 x 10 grid are shown in Figure A.4. The serial computation for Sapporo requires more than 7 times the computational time of the middle scale MtA. However using parallel computing the computational time for Sendai was just 26.6% of the serial process. For Sapporo, the computational time was 29.1% of its serial process. For that reason worth using parallel computing.

The total time for the SVM using a 10 x 10 grid presents the following equation:

$$\ln(CT_i) = 2.9709(MtA_i) + 5.8506 \tag{61}$$

$$CT_i = 337.17e^{2.9709 MtA_i} \tag{62}$$

Where:

MtA	is the Metropolitan area
CT_i	is the consumption time for the metropolitan area i
i	Defining 1 for Aomori, 2 for Sendai and 3 for Sapporo MtA-UPA

The equation (61) has an coefficient of determination $R^2 = 0.9532$ for the log-linear model. Figure A.5 shows the total time for SVM calculation using a 10x10 grid.

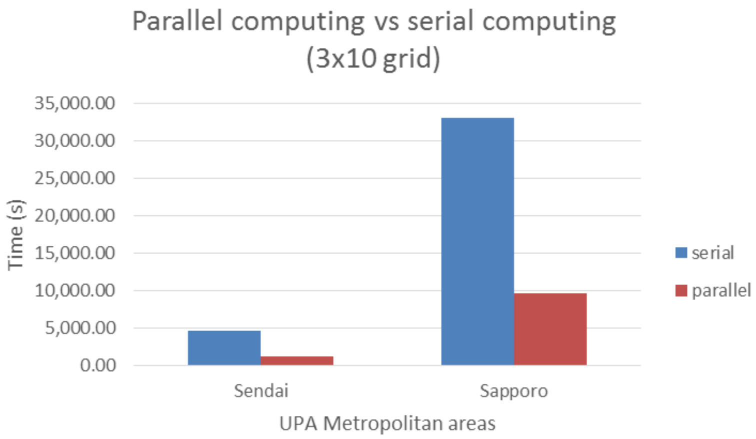


Figure A.4: Parallel computing vs serial computing

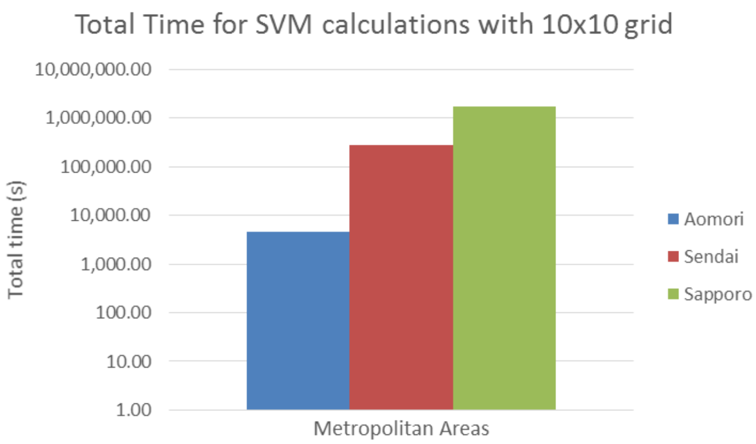


Figure A.5: Total time for SVM calculations with 10x10 grid

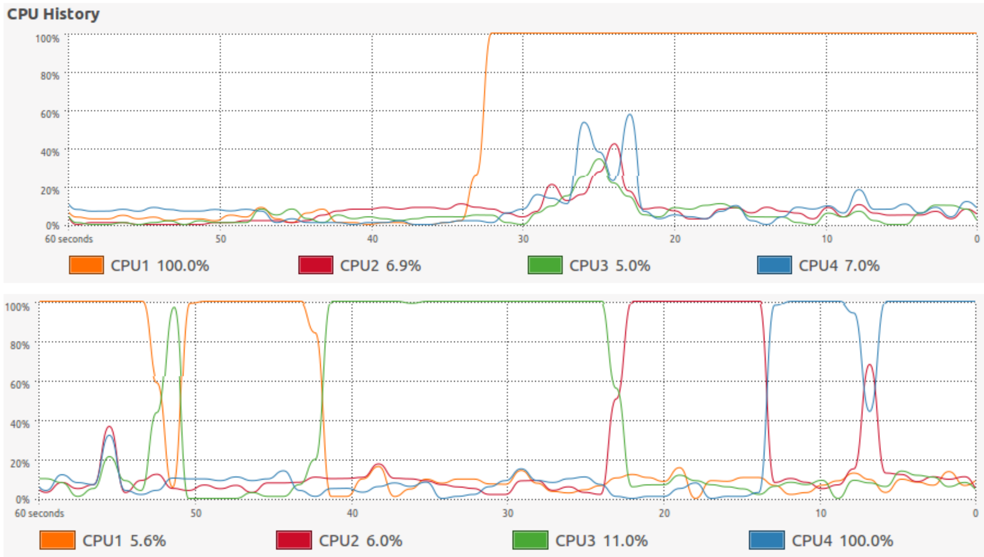


Figure A.6: Serial process behavior

Figure A.6 shows in the first part how one processor occupies 100% of capacity while the others remain with their tasks. In the second part, it is possible to see that just one processor is working using 100% and after a period of time the task is continued by another processor.

Figure A.7 shows the behavior of the parallel computing. The first row shows that all the processors start at the same time. In the second row it is possible to see that the processor called CPU2 is working just in 10% of its capacity, it means that a task given by one processor has finished. Third row, two processors are working using 100% of their capacity. Fourth row, according to this picture, when the time is 32 s just one processor remains working in 100%. Finally, in the last picture all the tasks have finished and CPU usage drops for all processors.

A.2.2 *Sendai MtA*

The information about the activity of the different processors using Parallel computing is shown in Figure A.8. According to this figure all the processors worked using over 95% of capacity in more than 95% of the processing time.

A.2.3 *Sapporo MtA*

Sample of processors activity using parallel computing is shown in Figure A.9. All the processors worked using over 95% of capacity.

Nevertheless some processors finished before the task, for this reason and to improve this situation it is important to continue researching on parallel computing.

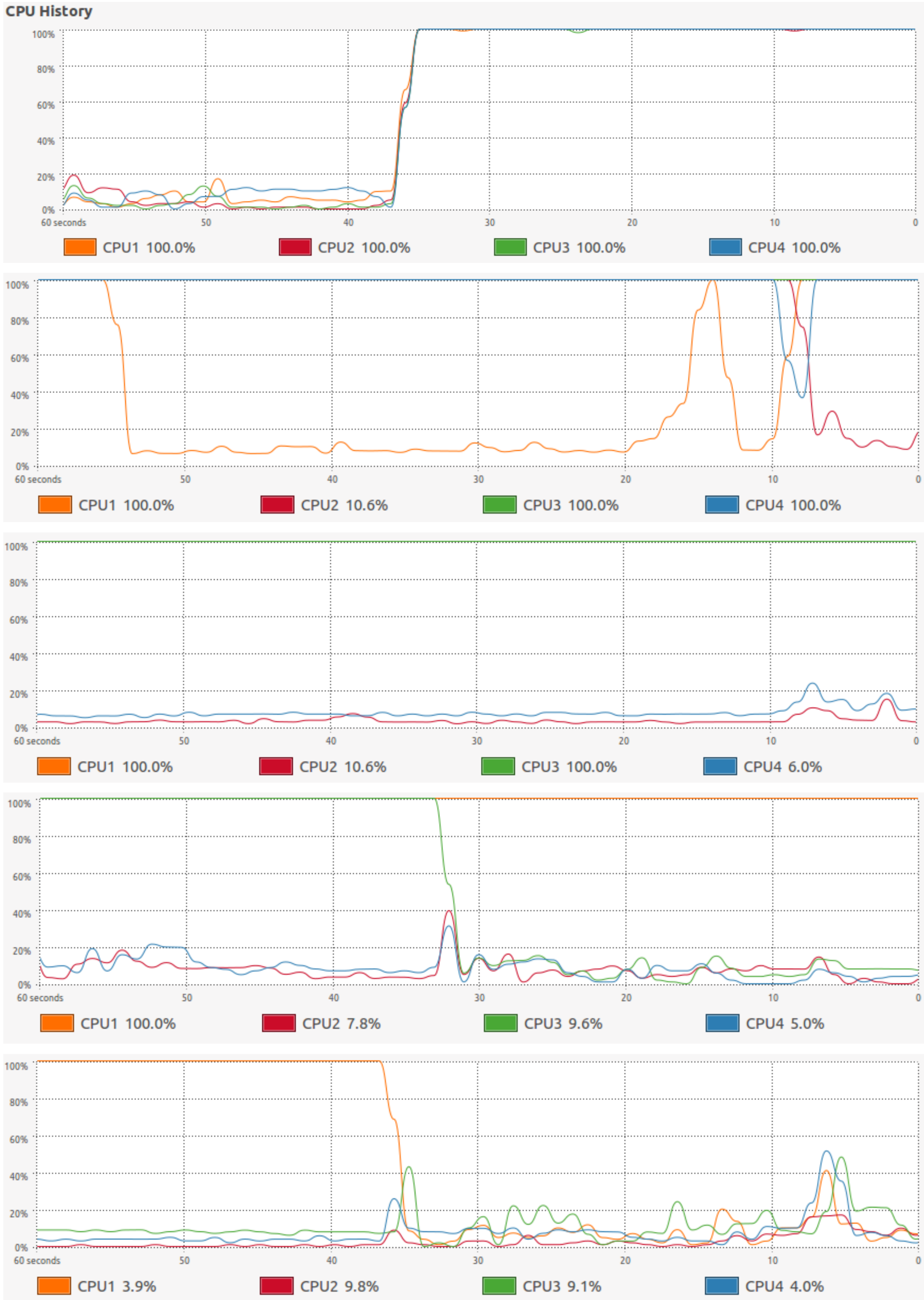


Figure A.7: Parallel process behavior

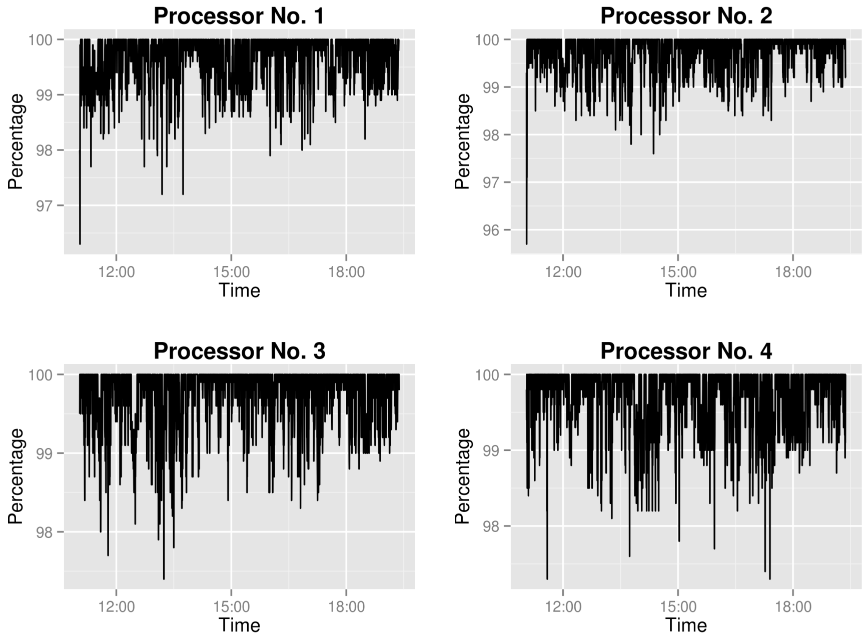


Figure A.8: Activity for all processor in the UPA of Sendai MtA (Sample)

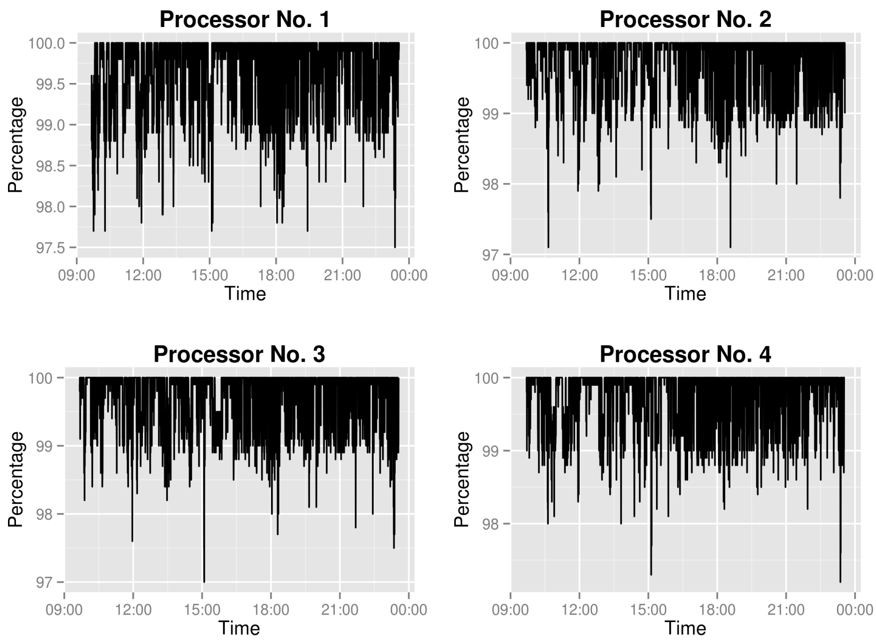


Figure A.9: Activity for all processor in the UPA of Sapporo MtA (Sample)

A.3 NOISE ANALYSIS

The unit to define the sound pressure level is the decibel (dB), table A.12 shows the different scales for this intensity. Under ideal conditions the humans talk between 60 - 65 dB, however the type 70dB are still in normal parameters[2].

0.0 dB is considered the threshold of hearing according to the internationally agreed upon quantity that decibel measurements are made. The information related with noise in the UPA of Aomori MtA is presented in tables A.13 and A.14. The values shown in these tables are the average of the sound intensity in the morning and afternoon. The values are in the normal range, for that reason it was not integrated as socio-economic factor. These information was the closest to the periods of time analyzed.

For the UPA of Sendai MtA, tables A.15 and A.16 show the information related with noise (dB), the value is the average of each type cross with line. Still the values are not enough high to think that affect the housing process.

Finally, for the UPA of Sapporo MtA, tables A.17 and A.18 present the information related with sound (dB), this value correspond to the average intensity level in the morning as well as afternoon. Although the values are higher than Sendai and Aomori there is not enough influence of the noise in the housing process. For that reason **noise** was not taken as a socio-economic factor for any MtA.

2 Decibel: http://www.sfu.ca/sonic-studio/handbook/Decibel.html

Table A.12: Noise information

type	Decibels(dB)
Threshold of hearing	0
Rustling leaves	20
Quiet whisper (3 feet)	30
Quiet home	40
Quiet street	50
Normal conversation	60
Inside car	70
Loud singing (3 feet)	75
Automobile (25 feet)	80
Motorcycle (30 feet)	88
Food blender (3 feet)	90
Subway (inside)	94
Diesel truck (30 feet)	100
Power mower (3 feet)	107
Pneumatic riveter (3 feet)	115
Chainsaw (3 feet)	117
Amplified Rock and Roll (6 feet)	120
Jet plane (100 feet)	130

Table A.13: Aomori's ROI noise (morning) (dB)

	type	3	4
year	lines		
2006	2	69	-
	4	69	-
2007	2	68.5	66.5
	6	72	-

Table A.14: Aomori's ROI noise (afternoon)(dB)

	category	3	4
year	lines		
2006	2	63	-
	4	64.5	-
2007	2	63	59.5
	6	68	-

Table A.15: Sendai's ROI noise (morning)(dB)

	type	1	3	4	5
year	lines				
2002	2	-	-	61	-
	4	67	69	70	-
	5	-	-	71	-
	6	-	79	69	73
2003	2	-	66	60	-
	4	66	67.5	66	67
	5	-	-	69	-
	6	-	78	69.5	70
2004	2	-	71	68.5	-
	4	61	69	67.25	65
	5	-	-	70	-
	6	-	73	-	70.5
2005	2	-	-	67.5	-
	4	58	70	68	65
	5	-	-	69	-
	6	-	79	66	75
2006	2	-	65	-	-
	4	67	68	69	63
	5	-	-	71	-
	6	-	73	-	63
2007	2	-	72	67.33	-
	4	-	67.67	70.5	67
	6	-	68	65	-

Table A.16: Sendai's ROI noise (afternoon)(dB)

year	category lines	1	3	4	5
2002	2	-	-	54	-
	4	64	67	65	-
	5	-	-	65.5	-
	6	-	76	64	70
2003	2	-	59	54	-
	4	64	66	60	62
	5	-	-	67	-
	6	-	76	63.5	65
2004	2	-	66	64.5	-
	4	59	67.33	63.25	58.67
	5	-	-	67	-
	6	-	69.5	-	66.5
2005	2	-	-	61.5	-
	4	56	66.5	64.33	62
	5	-	-	66	-
	6	-	77	61	72
2006	2	-	59	-	-
	4	64	64.67	65.25	57
	5	-	-	67	-
	6	-	69.5	-	56
2007	2	-	69	64	-
	4	-	63.67	65.75	63
	6	-	66	60	-

Table A.17: Sapporo's ROI noise (morning)(dB)

year	type lines	1	3	4	5	6
2002	2	-	71	-	-	66
	4	69.5	71.14	69.09	71	-
	5	-	-	72	-	-
	6	-	71	72	69.5	-
	8	-	-	69	-	-
2003	2	-	71	-	-	-
	4	72.5	72.43	70.45	68.71	-
	6	-	-	71.67	-	-
2004	2	-	-	69.6	-	-
	4	60	72	71.78	70.4	-
	5	-	-	71	71	-
	6	-	-	72	-	-
	7	-	70	-	-	-
2005	2	-	-	66	-	68
	4	-	71.67	69.14	67.81	-
	6	-	67	66.5	69	-
2006	2	-	-	71	-	-
	4	-	71.44	69.4	67.06	-
	6	-	65	72	71.5	-
2007	2	-	70	68.33	-	-
	4	-	70.2	68.83	66.27	-
	6	-	71	70	69.5	-

Table A.18: Sapporo's ROI noise (afternoon)(dB)

year	category lines	1	3	4	5	6
2002	2	-	65	-	-	61
	4	63.5	67.71	64.82	68	-
	5	-	-	69	-	-
	6	-	66	68	66	-
	8	-	-	66	-	-
2003	2	-	65	-	-	-
	4	65.5	69.43	66.09	63.43	-
	6	-	-	67.33	-	-
2004	2	-	-	64	-	-
	4	54	68.88	67.67	64	-
	5	-	-	65	67	-
	6	-	-	66	-	-
	7	-	69	-	-	-
2005	2	-	-	59.5	-	63
	4	-	68.33	63.14	62.29	-
	6	-	62	61	62.5	-
2006	2	-	-	64	-	-
	4	-	67.44	63.8	60.5	-
	6	-	59	67.5	66	-
2007	2	-	67	63.33	-	-
	4	-	66.4	64.5	59.91	-
	6	-	67	65	63.5	-

A.4 DISTANCE ANALYSIS THROUGH TRAFFIC NETWORKS

Distance analysis through traffic networks

Luis Carlos Manrique Ruiz
Yamamoto laboratory, University of Electro-communications

May 27, 2014

1 Introduction

City planners, architects among other professionals have been working on how to improve the network systems. This is useful because it is possible to create different scenarios to evaluate safety, travelling time, logistics, etc. The relation between the euclidean distance and the network distance have been studied since early 80's. The ratio between the network distance and the euclidean distance is defined as *Circuity* or *directness*. Newell [1] measured the distance between two points in a urban environment and discovered that there is a relation of 1.2 times the euclidean distance. In 1996, O'Sullivan [2] calculated circuity factors of 1.21 and 1.23 at various stations.

2 Distances

The experiments on this study are focused on finding a relationship between euclidean, driving and walking distances. Because it is necessary to construct all the network system (nodes and lines) to calculate the different distances, we have used Google system to calculate travel and walking distances. For this study we used their API [1].

An example of the API is given as follows: `http://maps.googleapis.com/maps/api/distancematrix/xml?origins=Vancouver+BC|Seattle&destinations=San+Francisco|Vancouver+BC&mode=bicycling&language=fr-FR&sensor=false&key=API_KEY`. Where the first part is related to the origin, second part is related to destination. The third part of this url is the transportation system, language, GPS sensor and user API_KEY.

Google Distance matrix offers to free users 100 elements per query, 100 elements per 10 seconds or 2,500 elements per 24 hour period. The result for the previous URL in XML format is shown as follows (Note that the transportation mode is bicycling):

```xml
<?xml version="1.0" encoding="UTF-8"?>
<DistanceMatrixResponse>
 <status>OK</status>
```

[1] The Google Distance Matrix API.
https://developers.google.com/maps/documentation/distancematrix/

```xml
 &lt;origin_address&gt;Vancouver, BC, Canada&lt;/origin_address&gt;
 &lt;origin_address&gt;Seattle, Etat de Washington, Etats-Unis&lt;/origin_address&gt;
 &lt;destination_address&gt;San Francisco, Californie, Etats-Unis&lt;/destination_address&gt;
 &lt;destination_address&gt;Victoria, BC, Canada&lt;/destination_address&gt;
 &lt;row&gt;
  &lt;element&gt;
   &lt;status&gt;OK&lt;/status&gt;
   &lt;duration&gt;
    &lt;value&gt;340110&lt;/value&gt;
    &lt;text&gt;3 jours 22 heures&lt;/text&gt;
   &lt;/duration&gt;
   &lt;distance&gt;
    &lt;value&gt;1734542&lt;/value&gt;
    &lt;text&gt;1 735 km&lt;/text&gt;
   &lt;/distance&gt;
  &lt;/element&gt;
  &lt;element&gt;
   &lt;status&gt;OK&lt;/status&gt;
   &lt;duration&gt;
    &lt;value&gt;24487&lt;/value&gt;
    &lt;text&gt;6 heures 48 minutes&lt;/text&gt;
   &lt;/duration&gt;
   &lt;distance&gt;
    &lt;value&gt;129324&lt;/value&gt;
    &lt;text&gt;129 km&lt;/text&gt;
   &lt;/distance&gt;
  &lt;/element&gt;
 &lt;/row&gt;
 &lt;row&gt;
  &lt;element&gt;
   &lt;status&gt;OK&lt;/status&gt;
   &lt;duration&gt;
    &lt;value&gt;288834&lt;/value&gt;
    &lt;text&gt;3 jours 8 heures&lt;/text&gt;
   &lt;/duration&gt;
   &lt;distance&gt;
    &lt;value&gt;1489604&lt;/value&gt;
    &lt;text&gt;1 490 km&lt;/text&gt;
   &lt;/distance&gt;
  &lt;/element&gt;
  &lt;element&gt;
   &lt;status&gt;OK&lt;/status&gt;
   &lt;duration&gt;
    &lt;value&gt;14388&lt;/value&gt;
    &lt;text&gt;4 heures 0 minutes&lt;/text&gt;
   &lt;/duration&gt;
   &lt;distance&gt;
    &lt;value&gt;135822&lt;/value&gt;
    &lt;text&gt;136 km&lt;/text&gt;
   &lt;/distance&gt;
  &lt;/element&gt;
 &lt;/row&gt;
&lt;/DistanceMatrixResponse&gt;
```

The previous result provides information related to traveling distance and

Table 1: Route studies (sample)

Route name	Haversine distance (km)	Driving		Walking	
		distance (km)	Time (s)	distance (km)	Time (s)
route_1_1	12.94	17.08	5,022	16.68	36,774
route_1_2	11.04	14.27	4,164	15.06	33,018
route_1_3	18.93	21.91	5,970	23.26	51,204
route_1_4	7.19	8.96	2,649	10.43	23,046
route_1_5	11.08	13.62	3,756	15.12	33,144
route_1_6	15.54	19.15	5,775	20.28	44,592
route_1_7	11.50	14.30	4,068	15.71	34,530
route_1_8	10.30	12.68	3,246	14.18	31,101
route_1_9	14.62	17.68	5,061	19.13	42,063
route_1_10	14.63	18.10	5,649	19.41	42,615

time. In section 6 an example of the google distance matrix application using Japanese information is provided in JSON format.

3 Experiments

We ran 100 experiments (routes), in Table 1 a sample is of the database shown. The routes are in located in the Urban Promotion Area (UPA) of Aomori Metropolitan Area (MtA), the table shows the haversine distance defined as the arc distance between two points and it is given in kilometres; once the points are closer to each other, the haversine distance tends to the euclidean distance. Later by using Google distance matrix API we calculate the driving distance (car) and travel time. Using the same application we calculate the walking distance and time also given in kilometres and seconds respectively.

According to Yamamoto et al[3], a walking distance can be described using euclidean distance. In their experiments they found that inside 3 special wards in Tokyo (Minato, toshima and Setagaya ku) present a slope same as 1.227, 1.250 and 1.204 respectively. In their experiments the calculation error was less than 3.4%. In our calculation by using Google's distance matrix results and comparing them by using GIS we found an accuracy of 98.68%, this is because Google is using Zenrin's information [2].

Figure 1 shows an example of a travel distance an the euclidean distance in Aomori MtA. In table 2 a sample of 4 travel routes are shown in the Aomori MtA. The bicycling route is similar with the walking route, for that reason it was not taken into account.

[2] Zenrin Corp. `http://www.zenrin.co.jp/`

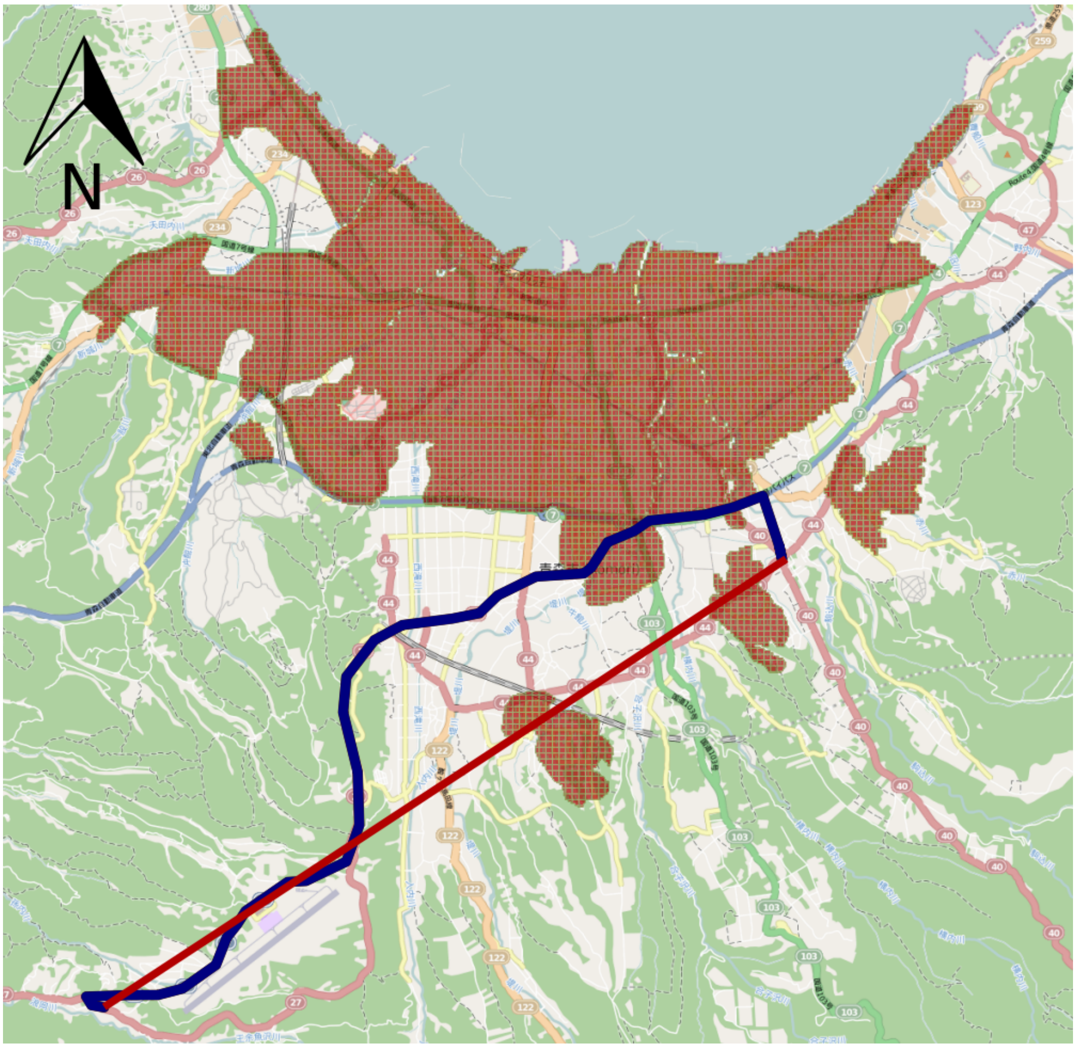


Figure 1: Route from A to B. (Blue: travel distance by car, red: euclidean distance)

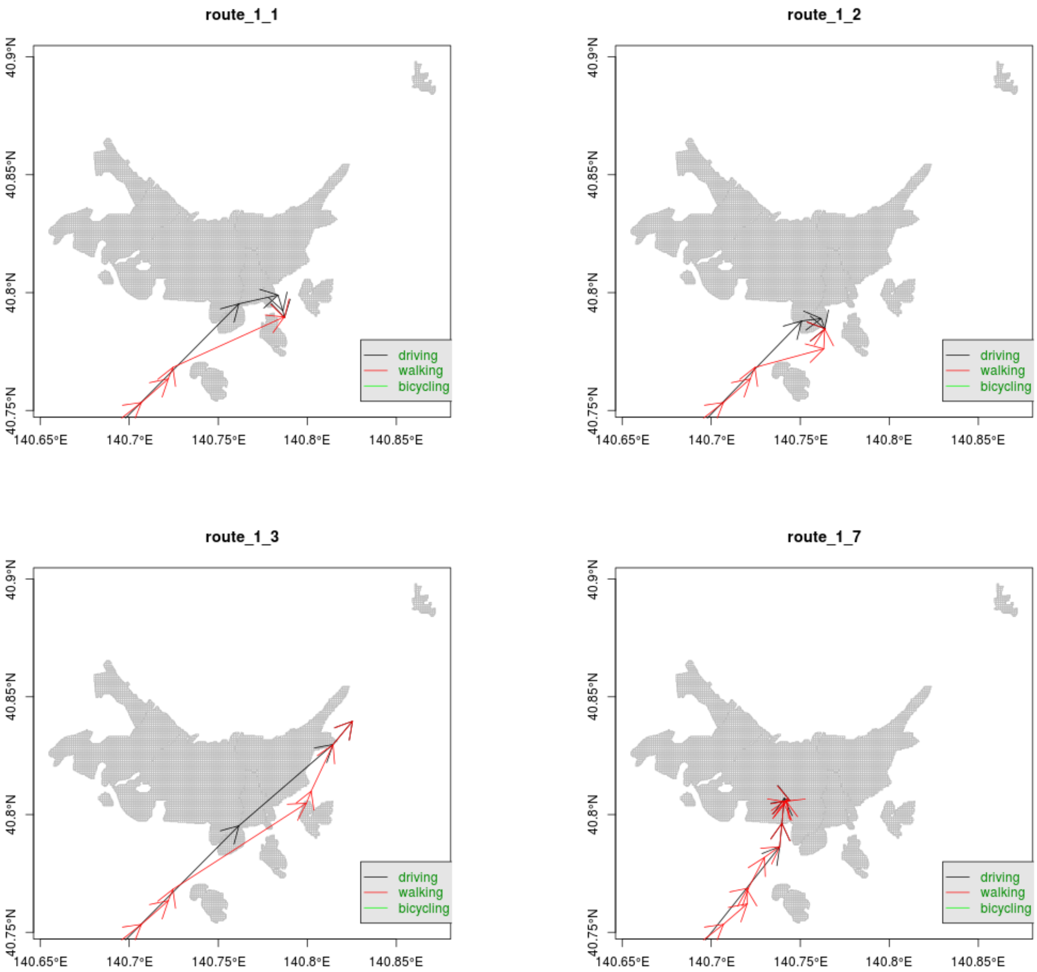


Table 2: travel distances

4 Results

For our 100 observations, we conclude that the haversine distance can be estimated by travel distance or walking distance. In table 3 the different models are shown, both are functions of the haversine distance. The first model is related with the travel distance by car, it presents a slope of 1.415 and a $R^2 = 0.968$. The model for walking distance presents a slope of 1.271 with and $R^2 = 0.991$. In both cases the F statistic and p-value are shown to justify the linear regression model.

Table 3: Regression Results

	Dependent variable:	
	T.dist.km	W.dist.km
	(1)	(2)
Haversine	1.415***	1.271***
	(0.026)	(0.012)
Observations	100	100
R^2	0.968	0.991
Adjusted R^2	0.967	0.991
Residual Std. Error (df = 99)	2.958	1.353
F Statistic (df = 1; 99)	2,955.771***	11,404.490***
Note:	*p<0.1; **p<0.05; ***p<0.01	

5 Conclusion

According to Yamamoto et al, the slope for every special ward in Tokyo is different; however the value is between 1.204 and 1.250. In our experiments, we demonstrated that walking and travel distance by car can be described by using the haversine distance. Toshima's walking model slope is similar with Aomori's MtA model slope (1.271)

References

[1] Gordon Frank Newell. *Traffic flow on transportation networks.* Number Monograph. 1980.

[2] Sean O'Sullivan and John Morrall. Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*, 1538(1):19–26, 1996.

[3] Kayoko Yamamoto and Hideharu Morishita. An evaluation of the locations of open spaces as regional refuge using geographical information systems (gis): A case study of tokyo. *Papers on environmental information science*, 12:125–130, nov 1998.

6 Appendix (Sample)

RESULTS (SAMPLE)

URL
http://maps.googleapis.com/maps/api/directions/json?origin=%E9%9D%92%E6%A3%AE%E7%9C%8C%E9%9D%92%E6%A3%AE%E5%B8%82%E5%A4%A7%E5%AD%97%E8%8D%92%E5%B7%9D&destination=%E9%9D%92%E6%A3%AE%E5%B8%82%E6%B8%AF%E7%94%BA1-19-4&mode=driving&units=metric&alternatives=false&sensor=false


```
{
   "routes" : [
      {
         "bounds" : {
            "northeast" : {
               "lat" : 40.8285051,
               "lng" : 140.7663264
            },
            "southwest" : {
               "lat" : 40.7829717,
               "lng" : 140.7373754
            }
         },
         "copyrights" : "Map data ©2014 Google, ZENRIN",
         "legs" : [
            {
               "distance" : {
                  "text" : "7.0 km",
                  "value" : 7023
               },
               "duration" : {
                  "text" : "19 mins",
                  "value" : 1161
               },
               "end_address" : "1 Chome-19-4 Minatomachi, Aomori, Japan",
               "end_location" : {
                  "lat" : 40.8285051,
                  "lng" : 140.7663264
               },
               "start_address" : "Arakawa, Aomori, Japan",
               "start_location" : {
                  "lat" : 40.7843996,
                  "lng" : 140.7386421
               },
               "steps" : [
                  {
                     "distance" : {
                        "text" : "41 m",
                        "value" : 41
                     },
                     "duration" : {
                        "text" : "1 min",
                        "value" : 30
                     },
                     "end_location" : {
                        "lat" : 40.7844715,
                        "lng" : 140.7381595
                     },
                     "html_instructions" : "Head \u003cb\u003ewest\u003c/b\u003e towards \u003cb\u003e
                      県道44号線\u003c/b\u003e",
                     "polyline" : {
                        "points" : "ou|wFo__zYM~A"
                     },
                     "start_location" : {
                        "lat" : 40.7843996,
                        "lng" : 140.7386421
                     },
                     "travel_mode" : "DRIVING"
                  },
                  {
                     "distance" : {
                        "text" : "0.2 km",
                        "value" : 170
                     },
                     "duration" : {
```

```
            "text" : "1 min",
            "value" : 43
         },
         "end_location" : {
            "lat" : 40.7829717,
            "lng" : 140.7377959
         },
         "html_instructions" : "Turn \u003cb\u003eleft\u003c/b\u003e onto \u003cb\u003e県道
          44号線\u003c/b\u003e",
         "maneuver" : "turn-left",
         "polyline" : {
            "points" : "}u|wFo|~yYvAR\\FtDj@"
         },
         "start_location" : {
            "lat" : 40.7844715,
            "lng" : 140.7381595
         },
         "travel_mode" : "DRIVING"
      },
      {
         "distance" : {
            "text" : "39 m",
            "value" : 39
         },
         "duration" : {
            "text" : "1 min",
            "value" : 46
         },
         "end_location" : {
            "lat" : 40.7831105,
            "lng" : 140.7373754
         },
         "html_instructions" : "Turn \u003cb\u003eright\u003c/b\u003e towards \u003cb\u003e
          県道44号線\u003c/b\u003e",
         "maneuver" : "turn-right",
         "polyline" : {
            "points" : "ql|wFgz~yYUfAEJ"
         },
         "start_location" : {
            "lat" : 40.7829717,
            "lng" : 140.7377959
         },
         "travel_mode" : "DRIVING"
      },
      {
         "distance" : {
            "text" : "0.1 km",
            "value" : 113
         },
         "duration" : {
            "text" : "1 min",
            "value" : 85
         },
         "end_location" : {
            "lat" : 40.7838791,
            "lng" : 140.7380231
         },
         "html_instructions" : "Turn \u003cb\u003eright\u003c/b\u003e towards \u003cb\u003e
          県道44号線\u003c/b\u003e",
         "maneuver" : "turn-right",
         "polyline" : {
            "points" : "mm|wFsw~yYqCcAA_@E["
         },
         "start_location" : {
            "lat" : 40.7831105,
            "lng" : 140.7373754
         },
         "travel_mode" : "DRIVING"
      },
      {
         "distance" : {
```

"text" : "0.3 km",
"value" : 280
},
"duration" : {
"text" : "1 min",
"value" : 55
},
"end_location" : {
"lat" : 40.7863696,
"lng" : 140.7384549
},
"html_instructions" : "Turn \u003cb\u003eleft\u003c/b\u003e onto \u003cb\u003e県道44号線\u003c/b\u003e",
"maneuver" : "turn-left",
"polyline" : {
"points" : "gr|wFs{~yY]GwASwBWcBSUAiCK"
},
"start_location" : {
"lat" : 40.7838791,
"lng" : 140.7380231
},
"travel_mode" : "DRIVING"
},
{
"distance" : {
"text" : "1.1 km",
"value" : 1106
},
"duration" : {
"text" : "2 mins",
"value" : 118
},
"end_location" : {
"lat" : 40.7962759,
"lng" : 140.7396816
},
"html_instructions" : "At \u003cb\u003e荒川字柴田（交差点）\u003c/b\u003e, continue onto \u003cb\u003e県道120号線\u003c/b\u003e",
"polyline" : {
"points" : "ya}wFi~~yYeGg@i@CSAsJi@QAYAkLw@yAIs@EyBMo@EiDWkIe@"
},
"start_location" : {
"lat" : 40.7863696,
"lng" : 140.7384549
},
"travel_mode" : "DRIVING"
},
{
"distance" : {
"text" : "3.0 km",
"value" : 3007
},
"duration" : {
"text" : "7 mins",
"value" : 420
},
"end_location" : {
"lat" : 40.8229402,
"lng" : 140.7450782
},
"html_instructions" : "Continue straight through \u003cb\u003e荒川字藤戸（交差点）\u003c/b\u003e to stay on \u003cb\u003e県道120号線\u003c/b\u003e",
"maneuver" : "straight",
"polyline" : {
"points" : "w__xF_f_zYeCQoCQo@GgDSeAIoEW_BKmDSuDUyF_@q@Gc@C}@EiBMqDSu@Gu@EqBMs@E}AKwAMSCsAOoAM_AMkD_@i@Go@@iBScI_Aw@Ko@GmFm@oDm@gCc@}EgAg@MeD}@e@MuBg@w@Sw@SqA]a@Ks@Ow@SWIw@Se@MkBo@eFu@"
},
"start_location" : {
"lat" : 40.7962759,

```
         "lng" : 140.7396816
      },
      "travel_mode" : "DRIVING"
   },
   {
      "distance" : {
         "text" : "1.8 km",
         "value" : 1793
      },
      "duration" : {
         "text" : "4 mins",
         "value" : 242
      },
      "end_location" : {
         "lat" : 40.8248202,
         "lng" : 140.7662243
      },
      "html_instructions" : "Turn \u003cb\u003eright\u003c/b\u003e at \u003cb\u003e国道柳
       町（交差点）\u003c/b\u003e onto \u003cb\u003e奥州街道(陸羽街道)/国道４号
       線\u003c/b\u003e",
      "maneuver" : "turn-right",
      "polyline" : {
         "points" : "kfdxFwg`zY?
        W@]OoBKiBAGEsACiAK_DUcFOcFSuGKcCEwAQ_FIuC[{HUuGOeFIgBEmAAw@I_BOuD]
        gE[cEGiBCc@McDCa@WaF"
      },
      "start_location" : {
         "lat" : 40.8229402,
         "lng" : 140.7450782
      },
      "travel_mode" : "DRIVING"
   },
   {
      "distance" : {
         "text" : "0.4 km",
         "value" : 398
      },
      "duration" : {
         "text" : "2 mins",
         "value" : 103
      },
      "end_location" : {
         "lat" : 40.8283384,
         "lng" : 140.7654515
      },
      "html_instructions" : "Turn \u003cb\u003eleft\u003c/b\u003e",
      "maneuver" : "turn-left",
      "polyline" : {
         "points" : "crdxF{kdzYcCTiAFw@H_AJmBNsBNWJa@Nu@Zg@L"
      },
      "start_location" : {
         "lat" : 40.8248202,
         "lng" : 140.7662243
      },
      "travel_mode" : "DRIVING"
   },
   {
      "distance" : {
         "text" : "76 m",
         "value" : 76
      },
      "duration" : {
         "text" : "1 min",
         "value" : 19
      },
      "end_location" : {
         "lat" : 40.8285051,
         "lng" : 140.7663264
      },
      "html_instructions" : "Turn \u003cb\u003eright\u003c/b\u003e\u003cdiv
       style=\"font-size:0.9em\"\u003eDestination will be on the left
       \u003c/div\u003e",
```

```
                  "maneuver" : "turn-right",
                  "polyline" : {
                     "points" : "chexFagdzYS{AMsA"
                  },
                  "start_location" : {
                     "lat" : 40.8283384,
                     "lng" : 140.7654515
                  },
                  "travel_mode" : "DRIVING"
               }
            ],
            "via_waypoint" : []
         }
      ],
      "overview_polyline" : {
         "points" : "ou|   wFo__zYM~AvARrEr@[rAqCcAA_@E[]GoEk@yBUiCKeGg@}@EeKk@eMy@
          mCO_TqAmN_Ai`@cCiUuAoGo@kFm@i@Go@@mLsAgBSmFm@oDm@gCc@}EgAmEkA{Cu@cFqA{DaAqC}@eFu@?
          W@]OoBMqBI}Ca@cKc@yNQ{E[uJq@qQa@sMYuGy@kK]sIWaFcCTaCPmDZsBNWJwAj@g@LS{AMsA"
      },
      "summary" : "県道 120 号線 and 奥州街道(陸羽街道)/国道 4 号線",
      "warnings" : [],
      "waypoint_order" : []
   }
],
"status" : "OK"
}
```

A.5 DOCUMENT FOR EXPERTS

Document for experts in city planning

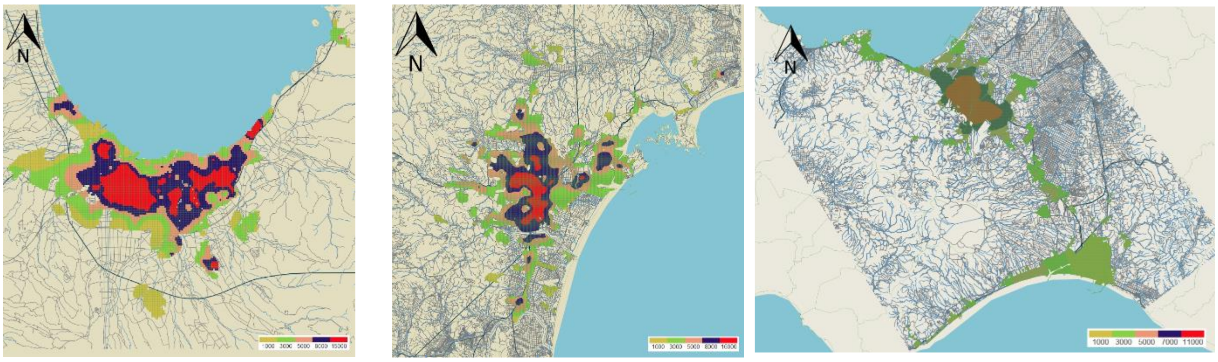
By: Luis Carlos Manrique Ruiz
Yamamoto laboratory, Department of Social Intelligence and Informatics
Graduate School of Information Systems
University of Electro-Communications

Objective: Verifying validity of an evaluation method for a compact city model.

Nowadays Japan faces problems related to depopulation and aging which are serious especially in local areas. For those reasons local governments apply some methods to deal with this situation. In Hokkaido and Tohoku regions, northern areas of Japan, some cities apply the compact city model to improve the social conditions; for instance, cities such as Aomori, Sendai and Sapporo. We focus on two main aspects of the compact city model: population density analysis, and the interaction of socio-economic factors with the land use. In this study, we extract patterns and indices from the Urbanization Promotion Area (UPA) in Aomori Metropolitan Area (MtA), because it has been working as a compact city model since 2000. Finally, we propose an evaluation method for a compact city model in the UPA of Aomori MtA and other MtAs of different sizes (Sendai and Sapporo MtAs).

Some scientists have been studying the compact city model from the point of view of land use and population density; however, there is no method that considers the relationship between these and socio-economic factors. In this study, we propose a new evaluation method which involves the three above-mentioned variables in order to assess the compact city model, and apply it in 3 different scales of metropolitan areas. The first part of the study was to identify the central, middle and outer areas by using population data in 3 different periods of time (1995, 2000 and 2005).

For this analysis, we apply a residual kriging model using a 100 m mesh in order to identify the population distribution (Figures 1, 2 and 3); the next part was related to compactness and entropy calculation from the point of view of the relationship between population and land use. In order to estimate compactness and entropy indexes we used Burton’s metrics and Shannon’s entropy respectively. Our experiments demonstrated that there is an inverse relation between these two indices (Figure 4).



Figures 1, 2, 3. Population density in Aomori, Sendai and Sapporo MtAs in 2005.

In the second part of this study, we predict land use through socio-economic factors in the three MtAs, however the socio-economic data was available just for the last period of time (2005). In order to estimate the land use we applied a Support Vector Machine (SVM) for classification; the information about socio-economic factors for the UPA in Aomori MtA is shown in Table 1.

Predicting land use using socio-economic factors by the SVM model, we found that the accuracy of the residential area was 93.3%. It was an appropriate indicator for the model, because the aim of the UPA is to promote the urban area. We used geographic information, land price and distance to socio-economic factors as model parameters, because they affect the land use. Using the last mentioned parameter data, we calculated the minimum number of combinations that describe all the residential area in the urban center (inner city), we called it ***number of vectors***.

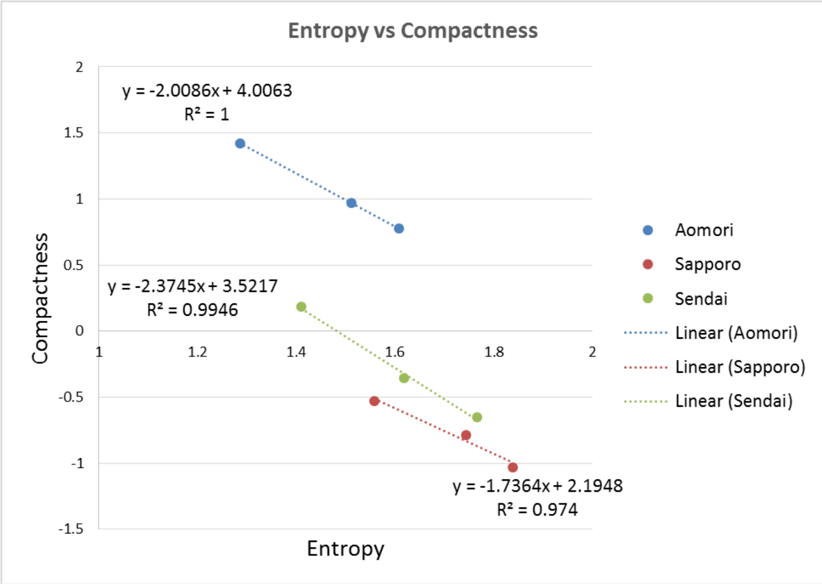


Figure 4. Relation between entropy and compactness

Table 1. Socio-economic factors for the UPA in Aomori MtA

Variable	**Additional information**	**Total**	**Type**
Bus stops		4,555	Num
Latitude, longitude		6,030	Num
Convenience stores	Circle K Sunkus, Sunkus, Ministop, Lawson	399	Num
Medical institutions	Ministry of health, labor and welfare, national hospital organizations, Social insurance, pension, sailors insurance, Hospital, Clinic, dental clinic, private medical institutions, welfare corporations.	1,522	Num
Parks		754	Num
Land price	By district	325	Num
Public facilities	Amusement parks, Other, National institutions, Local government, Welfare agency, Police agency, Fire station, School, Hospital, Post office, Welfare facilities.	3,216	Num / Cat
Supermarkets		85	Num
Train stations		9	Num
Land use data	Urbanization promotion area (UPA)	6,030	Cat

This number comes from a matrix containing the information related to mesh area and the distance interval between socio-economic factors.

The distance interval is defined as follows:
0 – 1 km = 1
1 – 3 km = 2
3 – 5 km = 3
5 – 10 km = 4
10 – 20 km = 5

Example:

Table 2. Example of a matrix containing distance to socio-economic factors

Id	Area	Distance A	Distance B	Distance C	…	Distance M
1	mesh 1	1	0	0	…	1
2	mesh 2	1	1	1	…	2
⋮	⋮	⋮	⋮	⋮	⋮	0
n	mesh n	2	1	**4**	…	5

Where the number highlighted in red represents that the socio-economic factor *C* is located between 5 and 8 km from the mesh *n*. In the previous example, there are *n* elements that represent all the combinations, and there are *M* socio-economic factors.

In order to continue evaluating the compact city model, we calculated a new index related to polygon compactness (PC), it is defined as follows:

$$\text{PC} = \frac{1}{core.value + dta}$$

Where:

core.value According to the compact city model and its application in Aomori MtA, 1 urban core should exists in the urban area.

dta Represents the number of urban areas detached from the urban core. We have defined a detached urban area, when its area is larger than 4.5% of the UPA. This percentage was defined by doing experiments using data of the three MtAs.

With the previous information we proceed to evaluate the compact city model, it is given by the following equation:

$$\text{Acid.value} = \text{Perc.UPA} * \log\left[Core.area + e^{\frac{Entropy*PC}{No.Vect}}\right] \tag{1}$$

Where:

Acid.value	Compactness value
Perc.UPA	Penalty factor in the equation, it represents the ratio between the urban core area and the total area of the Urbanization Promotion Area (UPA)
Core.area	Urban core area (ha)
Entropy	Shannon's entropy calculated for the UPA
PC	Polygon compactness (area which is larger than 4.5% of the UPA)
No.Vect	Number of vectors that define the core area by using socio-economic factors

Once the Acid value has been calculated for the three MtAs, we proceed to calculate the best value for this index.

Table 3. Best parameter values for indices

Parameter	Best value	Explanation
Perc.UPA	1	The urban core of the MtA is the same as the UPA
Entropy	0.08	Although Shannon's entropy threshold is defined as 1.0414, however the entropy tends to 0. The value 0.08 comes from the minimum entropy value calculated by Ramachandra. This value could be chosen by the researcher or city planner.
PC	1	The urban core is same as the UPA and hence detached large urban areas are equal to 0.
No.Vect	40	The number of combinations explains the residential area for the compact city model in the UPA of Aomori MtA.

Table 4. Parameters for acid value calculation

MtA	Percentage of UPA	area(ha)	entropy	PC	No. vectors
Aomori	87%	5,261	1.286	1.00	40
Sendai	28%	8,746	1.411	0.33	54
Sapporo	32%	23,625	1.559	0.20	72

The results for the acid test by using the population data in 2005, land use and socio-economic factors are:

Table 5. Results of compact city model evaluation

MtA	Acid value	Acid value (Best values)	Acid value / best values
Aomori	7.48	8.57	87.3%
Sendai	2.54	9.08	28.0%
Sapporo	3.22	10.07	32.0%

According to Aomori's data of 2005, the ratio between the population of the UPA and the metropolitan area is 92.9%. This value is similar to the ratio between the acid value and the acid value with best values for Aomori MtA (87.3%); results are shown in Table 5. We conclude that Aomori MtA is a good example for the compact city model; however, it needs to reduce the size of detached urban areas in the UPA, and continue promoting the residential area.

The basis of the equation (1) relies on the importance of the clear boundaries. It is necessary that residents should concentrate in the urban core, because that will help to improve the land use and the socio-economic factors which will reduce transportation costs, electric power transmission and so on. The **acid value/best value** could be interpreted as percent complete of the compact city model.

Generalization

Because Japan has strict regulations in terms of the land use, and it also develops large quantity of data, it is difficult for other countries to apply the evaluate method which we propose in this study to their metropolitan areas. However, it is possible to use Densely Inhabited Districts (DID) in order to identify the core of metropolitan areas and detached urban areas in other countries. It is necessary to collect the information related to the land use in order to calculate Shannon's entropy, and socio-economic information to calculate the number of vectors to describe the residential area.

The generalized model to evaluate a compact city model can be written as:

$$\mathrm{Acid.value} = \sum_{i=1}^{n} \left[\mathrm{Perc.MtA} * \log \left[Core.area + e^{\frac{Entropy*PC}{No.Vect}} \right] \right]_i \qquad (2)$$

Where:

n is the total number of urban cores in the MtA.

Note: There is a limitation in the data acquisition, because there are available satellite data and printed maps in other countries; however, there is not enough detailed information related to socio-economic factors. In order to evaluate compact city model, it is required to collect information about the parameters involved in this study to promote the residential area in the urban area.

Adding a summation in Equation (1), it is possible to evaluate different scales of MtAs; however, it is necessary to evaluate separately the urban cores (central and regional ones). For instance, Sapporo MtA has 3 urban cores respectively in Sapporo, Otaru and Tomakomai cities; so as to evaluate properly the compact city model in this MtA, it is necessary to calculate the acid value for the three areas which have each urban core separately. Once the values in each of the three areas are calculated, the new acid value for Sapporo MtA will rise. Finally, based on the equation (2), evaluating the three areas respectively, it will be possible to decrease errors.

A.6 CODE

In order to develop this research, R software and Tilemill were used to manipulate GIS data. In this part the code will be added, it contains not only the main results but also another experiments developed to calculate the best model.

CONSTRUCTING POPULATION

```
#written in UTF-8 encoding

library(maptools)
llCRS<-CRS("+proj=longlat +ellps=WGS84")

#maindir<-"~/Documents/phd/surrounding
sendai/" #linux
maindir<-"C:/Users/Kawai/Documents/carlos/P
hD/surrounding sendai"
setwd(maindir)
dir()
tmp<-list.dirs()
tmpd<-tmp[grep("y17",tmp )]
tmpd

#extracting all the documents from zip
files
for(i in 1:length(tmpd))
{
#  i=1
setwd(tmpd[i])
getwd()
 cat(paste(tmpd[i]), "/n")
print(dir())
cat("-------------------","/n")
tmp<-dir()
tmp<-tmp[grep("zip",tmp)]
for(j in 1:length(tmp)) unzip(tmp[j],
overwrite = TRUE)
setwd(maindir)
}

###########################################
#starting folders
tmpd
setwd(maindir)

for(k in 1:length(tmpd))
{
#       k=2
        setwd(tmpd[k])
        getwd()
        tmp<-dir()
        tmp<-tmp[grep("tbl", tmp)]
        tmp<-tmp[grep(".txt", tmp)]

        #options(encoding="Shift-JIS")
        for(i in 1:length(tmp))
        {
        #  i=1
         tmpn<-paste("file",i,sep="")
         assign(tmpn, read.table(tmp[i],
sep=",", skip=2, colClasses = "character"))

         #temporal file to extract the names
         assign(paste("tmpfile") ,
read.table(tmp[i], sep=",",colClasses =
"character"))
         newf<-get(tmpn)
         names(newf) <-
as.character(c(tmpfile[1,c(1:5)],tmpfile[2,
c(6:dim(tmpfile)[2])]))
         head(newf)
         assign(tmpn, newf)
        }
        head(file1)
        rm("tmpfile")
        ls()
        tmp<-ls()[grep("file", ls())]
        tmp<-setdiff(tmp,
tmp[grep(".csv",tmp)])

        for(j in 1:length(tmp))
        {
        write.csv(get(tmp[j]),
paste("file",j,".csv",sep=""), row.names=F)
        }

        for(i in 1:length(tmp))
        {
         assign(tmp[i], get(tmp[i])[,-
which(names(get(tmp[i]))=="HTKSYU")])
        }

        newf<-file1
        for(i in 2:length(tmp))
        {
        newf<-merge(newf, get(tmp[i]))
        cat(print(dim(newf)),"/n")
        }
        tmpn1<-substr(tmpd[k], 3,
nchar(tmpd[k])-4)
        tmpn1<-paste(gsub(" ","", tmpn1),
".csv", sep="")
        write.csv(newf, tmpn1, row.names=F)

setwd(maindir)
}

#---- created files
###########################################
#---- starting
#options(encoding="UTF-8")
getwd()
setwd(maindir)

for(k in 1:length(tmpd))
{
#k=1
        setwd(tmpd[k])
 getwd()
        tmp<-dir()
        tmp<-tmp[grep(".shp", tmp)]
        #tmp<-tmp[grep("utf",tmp)]

        tmpn1<-substr(tmpd[k], 3,
nchar(tmpd[k])-4)
        tmpn1<-gsub(" ","", tmpn1)

        assign(tmpn1,
readShapePoly(tmp,proj4string=llCRS))
#}
 #plot(get(tmpn1), axes=T)
        tmpshp<-tmpn1 #shapefile
        tmpdb <-paste(tmpn1,"_file",sep="")
#database file

        #dir()

        #options(encoding="Shift-JIS")
        #wataricho_file<-
read.csv("wataricho.csv",header=T,colClasse
```

```
s = "character")

assign(tmpdb,read.csv(paste(tmpn1,".csv",se
p=""),header=T,colClasses = "character") )
        #options(encoding="UTF-8")

        #ls()
        #plot(get(tmpshp))
        #length(get(tmpshp)) #86
        #str(get(tmpshp))

        db<-get(tmpshp)@data
                # area is given in m2
        #head(db)
        #n<-10
        #get(tmpshp)@polygons[[n]]
        #slot(get(tmpshp)@polygons[[n]],
"ID" )
        #slot(get(tmpshp)@polygons[[n]],
"labpt" ) #[1] 140.90790  37.99702
        #slot(get(tmpshp)@polygons[[n]],
"area" )
        #head(db)
        #dim(db)

        tmp<-get(tmpdb)
        baby<-"0.4"
        young<-"15.19"
        mature<-"35.39"
        oldp<-"75"
        #names(tmp)
        #grep(young, names(tmp))

        # subsets
        # population
        pop_tmp<-grep("総数",names(tmp))
        cont<-grep("総数",names(tmp))
        pop_tmp<-tmp[,pop_tmp]
        #head(pop_tmp)
        pop_tmp<-pop_tmp[ ,-match(
        c("総数.年齢.不詳.含む" ,"総数15歳未満","
総数15.64歳","総数65歳以上","総数65.74歳","総数
の総年齢", "男の総数.年齢.不詳.含む","女の総数.年
齢.不詳.含む"),
        names(pop_tmp))]

        #names(pop_tmp)

        tmpbaby<-grep(baby, names(pop_tmp))
[1]
        tmpyoun<-grep(young, names(pop_tmp))
[1]
        tmpmat<-grep(mature, names(pop_tmp))
[1]
        tmpold<-grep(oldp, names(pop_tmp))[1]

        for(i in 1:dim(pop_tmp)[2])
        {
        pop_tmp[,i]<-as.numeric(pop_tmp[,i])
        }

        tbab<-
apply(pop_tmp[,c(tmpbaby:tmpyoun)], 1,sum,
na.rm=TRUE) #baby ~ young <- children
people
        tyoung<-apply(pop_tmp[,c(
(tmpyoun+1):tmpmat)], 1,sum, na.rm=TRUE)
#young+1 ~ mature  <- young
        tmat<-apply(pop_tmp[,c( (tmpmat+1):
(tmpold-1))], 1,sum, na.rm=TRUE) #mature+1 ~
old <-mature
        told<-pop_tmp[,tmpold]
        #head(pop_tmp)
        pop_tmp1<-cbind(pop_tmp, tbab,
tyoung, tmat, told)
        #getwd()
        write.csv(pop_tmp1, "poptmp.csv",
row.names=F)

        #men
        men_tmp<-grep("男",names(tmp))
        cont<-c(cont, men_tmp)
        men_tmp<-tmp[,men_tmp]
        #head(men_tmp)

        men_tmp<-men_tmp[,-match( c("男.","男
の総数.年齢.不詳.含む","男15歳未満","男15.64歳",
"男65歳以上","男65.74歳","男の総年齢"),
names(men_tmp))]
        #names(men_tmp)

        tmpbaby<-grep(baby, names(men_tmp))
[1]
        tmpyoun<-grep(young, names(men_tmp))
[1]
        tmpmat<-grep(mature, names(men_tmp))
[1]
        tmpold<-grep(oldp, names(men_tmp))[1]

        for(i in 1:dim(men_tmp)[2])
        {
        men_tmp[,i]<-as.numeric(men_tmp[,i])
        }

        tbab<-
apply(men_tmp[,c(tmpbaby:tmpyoun)], 1,sum,
na.rm=TRUE) #baby ~ young <- children
people
        tyoung<-apply(men_tmp[,c(
(tmpyoun+1):tmpmat)], 1,sum, na.rm=TRUE)
#young+1 ~ mature  <- young
        tmat<-apply(men_tmp[,c( (tmpmat+1):
(tmpold-1))], 1,sum, na.rm=TRUE) #mature+1 ~
old <-mature
        told<-men_tmp[,tmpold]
        #head(men_tmp)
        men_tmp1<-cbind(men_tmp, tbab,
tyoung, tmat, told)
        #getwd()
        write.csv(men_tmp1, "mentmp.csv",
row.names=F)


        #woman
        wmen_tmp<-grep("女",names(tmp))
        cont<-c(cont,wmen_tmp)
        wmen_tmp<-tmp[,wmen_tmp]
        #names(wmen_tmp)

        wmen_tmp<-wmen_tmp[,-match( c("女","女
```

```
の総数.年齢.不詳.含む","女15歳未満","女15.64歳
","女65歳以上","女65.74歳","女の総年齢"),
names(wmen_tmp))]
	#names(wmen_tmp)

	tmpbaby<-grep(baby, names(wmen_tmp))
[1]
	tmpyoun<-grep(young, names(wmen_tmp))
[1]
	tmpmat<-grep(mature, names(wmen_tmp))
[1]
	tmpold<-grep(oldp, names(wmen_tmp))
[1]

	for(i in 1:dim(wmen_tmp)[2])
	{
	wmen_tmp[,i]<-
as.numeric(wmen_tmp[,i])
	}

	tbab<-
apply(wmen_tmp[,c(tmpbaby:tmpyoun)], 1,sum,
na.rm=TRUE) #baby ~ young <- children
people
	tyoung<-apply(wmen_tmp[,c(
(tmpyoun+1):tmpmat)], 1,sum, na.rm=TRUE)
#young+1 ~ mature  <- young
	tmat<-apply(wmen_tmp[,c( (tmpmat+1):
(tmpold-1))], 1,sum, na.rm=TRUE) #mature+1 ~
old <-mature
	told<-wmen_tmp[,tmpold]
	#head(wmen_tmp)
	wmen_tmp1<-cbind(wmen_tmp, tbab,
tyoung, tmat, told)
	#getwd()
	write.csv(wmen_tmp1, "wmentmp.csv",
row.names=F)

	#-----------------
	#householders
	#names(tmp)
	#head(tmp[,-cont])
	hh_tmp<-grep("世帯",names(tmp))
	cont<-c(cont,hh_tmp)
	#-------------
	cont<-unique(cont)
	#Creating sumarized file
	fd<-tmp[,-cont]
	#head(fd)
	 #pop
	fd<-cbind(fd, pop_tmp1[,c(
(dim(pop_tmp1)[2]-3) :dim(pop_tmp1)[2])])
	names(fd)[ (dim(fd)[2]-3):dim(fd)[2]
]<-paste("総数",
c(baby,young,mature,oldp),sep="")
	#head(fd)
	 #men
	fd<-cbind(fd, men_tmp1[,c(
(dim(men_tmp1)[2]-3) :dim(men_tmp1)[2])])
	names(fd)[ (dim(fd)[2]-3):dim(fd)[2]
]<-paste("男",
c(baby,young,mature,oldp),sep="")
```

```
	#head(fd)
	 #women
	fd<-cbind(fd, wmen_tmp1[,c(
(dim(wmen_tmp1)[2]-3) :dim(wmen_tmp1)[2])])
	names(fd)[ (dim(fd)[2]-3):dim(fd)[2]
]<-paste("女",
c(baby,young,mature,oldp),sep="")
	fd<-replace(fd, is.na(fd), 0)
#replacing all na values with 0
	#head(fd)


#------------------------------------------

#------------------------------------------
	fd$lat<-mat.or.vec(dim(fd)[1],1)
	fd$lon<-mat.or.vec(dim(fd)[1],1)
	fd$area<-mat.or.vec(dim(fd)[1],1)
	fd$perimeter<-mat.or.vec(dim(fd)
[1],1)

	#head(fd)
	#head(db)

	for(i in 1:dim(fd)[1])
	{
		tmp<-which(db$KEY_CODE ==
fd$KEY_CODE[i])
		if(length(tmp)>0)
		{
		fd$lon[i]<-db[tmp, "X_CODE"]
		fd$lat[i]<-db[tmp, "Y_CODE"]
		fd$area[i]<-db[tmp, "AREA"]
		fd$perimeter[i]<-db[tmp,
"PERIMETER"]
		}

		if(length(tmp)==0)
		{
		fd$lon[i]<-0
		fd$lat[i]<-0
		fd$area[i]<-0
		fd$perimeter[i]<-0
		}
	}

 head(fd)
	#fd
	fd$density<-(apply(fd[,grep("総数
",names(fd) )] ,1,sum)/fd$area)*(1000^2)
#density 人/km2
	fd$density<-replace(fd$density,
is.infinite(fd$density), 0) #replacing all
na values with 0
	#fd

	#writing file
	write.csv(fd, "finaldata.csv",
row.names=FALSE)
	setwd(maindir)
}
```

COMPACTNESS VS ENTROPY

```
library(ggplot2)

maindir<-"C:/Users/kuri73/Documents/carlos/PhD/t
hesis/Carlos/files/general"
setwd(maindir)

dir()

compact<-read.csv("compactness.csv", header=T)
compact

entr<-read.csv("entropy.csv", header=T)
names(entr)<-c("city", "year","Entropy", "Code")
entr

newdb<-merge(compact, entr)
newdb

aom<-subset(newdb, newdb$city=="Aomori")
aomlm<-lm(compact~Entropy, data=aom)
summary(aomlm)
summary(aov(compact~Entropy, data=aom))

sap<-subset(newdb, newdb$city=="Sapporo")
saplm<-lm(compact~Entropy, data=sap)
summary(saplm)

send<-subset(newdb, newdb$city=="Sendai")
sendlm<-lm(compact~Entropy, data=send)
summary(sendlm)

p<-ggplot(aom, aes(Entropy, compact)) +
geom_point()


ggplotRegression <- function (fit) {
       require(ggplot2)
       ggplot(fit$model, aes_string(x =
names(fit$model)[2], y = names(fit$model)[1])) +
        geom_point() +
        stat_smooth(method = "lm", col = "red") +
        opts(title = paste("Adj R2 =
",signif(summary(fit)$adj.r.squared, 5),
          "; Intercept =",signif(fit$coef[[1]],5
),
          "; Slope =",signif(fit$coef[[2]], 5),
          "; P =",signif(summary(fit)$coef[2,4],
5)))
}

ggplotRegression(sendlm)
```

PLOTTING LEGENDS on TileMILL

```
<div class='my-legend'>
<div class='legend-title'>Metropolitan areas
(UPA)</div>
<div class='legend-scale'>
 <ul class='legend-labels'>
  <!-- This legend uses Unicode box-drawing
characters to approxmate line styles. -->
  <li><span style='background:#d1bf49'></span>
1000 </li>
  <li><span style='background:#88d147'></span>
3000 </li>
  <li><span style='background:#ec9876'></span>
5000 </li>
  <li><span style='background:#2a1569'></span>
8000 </li>
  <li><span style='background:#f40a1b'></span>
15000 </li>
 </ul>
</div>
</div>

<style type='text/css'>
 .my-legend .legend-title {
  text-align: left;
  margin-bottom: 8px;
  font-weight: bold;
  font-size: 150%;
  }
 .my-legend .legend-scale ul {
  margin: 0;
  padding: 0;
  float: left;
  list-style: none;
  }
 .my-legend .legend-scale ul li {
  display: block;
  float: left;
  width: 50px;
  margin-bottom: 6px;
  text-align: center;
  font-size: 120%;
  list-style: none;
  }
 .my-legend ul.legend-labels li span {
  display: block;
  float: left;
  height: 15px;
  width: 50px;
  }
 .my-legend .legend-source {
  font-size: 70%;
  color: #999;
  clear: both;
  }
 .my-legend a {
  color: #777;
  }
</style>
```

PLOTTING MAPS

```
@n:0.8;
Map {
 background-color: #85c5d3;
}

#countries {
 ::outline {
  line-color: #85c5d3;
  line-width: 2;
  line-join: round;
 }
 polygon-fill: #ebe9de;
}

#AomPref {
 line-color:#594;
 line-width:0.5;
 polygon-opacity:1;
 polygon-fill:#ebe9de;
}

#aomrailwaysroi {
 line-width:2;
 line-color:#083649;
}

#aomroadsroi {
 line-width:1;
 line-color:#6e7377;
}

#aomwaterwaysroi {
 line-width:1;
 line-color:#4687b9;
}

@n: 0.8;

#km_popdens12 {
 line-color:#594;
 line-width:0.0;
 polygon-opacity:1;
 polygon-fill:#71ad53;
}

/* Standard Deviation */

#km_popdens12[se<=0.325655] {
 line-color:#594;
 line-width:0.0;
 polygon-opacity:@n;
 polygon-fill:#f0baba;
}

#km_popdens12[se>0.325655][se<=0.37642]{
 line-color:#594;
 line-width:0.0;
 polygon-opacity:@n;
 polygon-fill:#e97575;
}

#km_popdens12[se>0.37642][se<=0.43620]{
 line-color:#594;
 line-width:0.0;
 polygon-opacity:@n;
 polygon-fill:#e54949;
}
```

```
#km_popdens12[se>0.43620][se<=0.52791]{
 line-color:#594;
 line-width:0.0;
 polygon-opacity:@n;
 polygon-fill:#b61f1f;
}

#km_popdens12[se>0.52791][se<=1]{
 line-color:#594;
 line-width:0.0;
 polygon-opacity:@n;
 polygon-fill:#961219;
}

/* Population density*/
/*
#km_popdens12[density>1][density<=1000] {
 line-color:#594;
 line-width:0.5;
 polygon-opacity:1;
 polygon-fill:#d1bf49;
}

#km_popdens12[density>1000][density<=3000] {
 line-color:#594;
 line-width:0.5;
 polygon-opacity:1;
 polygon-fill:#88d147;
}

#km_popdens12[density>3000][density<=5000] {
 line-color:#594;
 line-width:0.5;
 polygon-opacity:1;
 polygon-fill:#ec9876;
}

#km_popdens12[density>5000][density<=8000] {
 line-color:#594;
 line-width:0.5;
 polygon-opacity:1;
 polygon-fill:#2a1569;
}

#km_popdens12[density>8000][density<=557542] {
 line-color:#594;
 line-width:0.5;
 polygon-opacity:1;
 polygon-fill:#f40a1b;
}*/
```

SVM MODEL FOR AOMORI MtA UPA and AUC calculation

```
# TODO: Add comment
#
# Author: Carlos
##################################################
rm(list=ls(all=TRUE))
setwd("C:/Users/kuri73/Documents/carlos/PhD/SVM")
#dir()
load("svm.Rdata")
ls()

setwd("C:/Users/kuri73/Documents/carlos/PhD/experiments/third part")
dir()
load("rvm.Rdata")
load("aom_map.Rdata")
#-----------------------------------
#install.packages("xtable")
library(kernlab)
library(mail)
library(rgl)
library(sp)
library(shapefiles)
library(xtable)
library(e1071)
library(caret)
#-----------------------------------
#Notes
#Support Vector Machine
#If the value of sigma is increased the training error decreases.
#If the cost is high and the sigma value is also high the parameter error decreases
#If I increase the cross value the error stills same.

#-------------------------------------------------
#when data come from rvm, they are not scaled.
tmp<-scale(trainData[,c(4:13)])
trainData<-cbind(trainData[,(1:3)], tmp, trainData[,(14:15)]) #training data scaled

tmp<-scale(testData[,c(4:13)])
testData<-cbind(testData[,(1:3)], tmp, testData[,(14:15)]) #testing data scaled

tmp<-scale(compdt[,c(4:13)])
compdt<-cbind(compdt[,(1:3)], tmp, compdt[,(14:15)]) #complete database

#-------------------------------------------------
#************* Working with the svm function *******************
dim(trainData) #[1] 4248   15
dim(testData) #[1] 1783   15
dim(compdt)  #[1] 6031   15

myformula<-土地利用種~.
lev1<-levels(trainData$土地利用種)
#changing land use types
levels(compdt$土地利用種)<-levels(trainData$土地利用種)<-levels(testData$土地利用種)<-c("a","b","c","d","e","f","A","B","E","F")
rbind(lev1, levels(compdt$土地利用種))

svm.m1 <- ksvm(myformula, data=trainData, kernel="rbfdot", kpar=list(sigma=0.05), C=5, cross=3)
svm.m1

#Training error : 0.2455
#Cross validation error : 0.289
svm.m1p <- predict(svm.m1, testData[,-3])
table(svm.m1p)
table(testData[,3])
table(svm.m1p, testData[,3])
sum(svm.m1p==testData[,3])/length(testData[,3])
# [1] 0.7178

svm.m2 <- ksvm(myformula, data=trainData, kernel="rbfdot", kpar=list(sigma=0.01), C=1000, cross=3)
#Training error : 0.194
#Cross validation error : 0.284

svm.m2p <- predict(svm.m2, testData[,-3])
table(svm.m2p)
table(testData[,3])
table(svm.m2p, testData[,3])
sum(svm.m2p==testData[,3])/length(testData[,3])
#[1] 0.74

#cost C = 2000
#Training error : 0.168785
#Cross validation error : 0.285075

#cost C = 3000
#Training error : 0.157486
#Cross validation error : 0.282486

#------------------------------------
#starting simulations

#Ct<-seq(1,75,1) #cost 30
#Cr<-seq(2,50,1) #cross 30

Ct<-seq(10,100,10) #cost 30
Cr<-seq(10,100,10) #cross 30

length(seq(10,100,10)) + length(seq(1,75,1))
length(seq(10,100,10)) + length(seq(2,50,1))

ptm<-proc.time()
for(i in Ct)
{
	for(j in Cr)
	{
		#if(j==1)
		#{
	#			j<-j+10
	#		} else if(j!=1){
		#	i<-0.1; j<-0.99

	assign(paste("svm_",i,"_",j,sep=""),
ksvm(myformula, data=trainData, kernel="rbfdot", kpar="automatic", C=i, cross=j) )
			cat(i, j, "\n")
		}
	}
	save.image("svm.Rdata")
}
ptmf<-proc.time()-ptm

tmp<-ls()[grep("svm_",ls())]
svm.m1
svm.m2

simdb<-NULL #simulation database
for (i in tmp)
{
	new<-get(i)
	simdb<-rbind(simdb,
	c(unlist(strsplit(i,"_"))[c(2,3)], #cost and cross
	  new@kernelf@kpar$sigma, #sigma parameter
		new@nSV, # number of support vectors
		new@error, #Training error
		new@cross) #Cross valiation error
		)
	print(dim(simdb))
```

```
}
simdb<-as.data.frame(simdb)
names(simdb)<-c("cost",
"cross","sigma","nsv","trer","crer")
simdb$cost<-as.numeric(as.character(simdb$cost))
simdb$cross<-as.numeric(as.character(simdb$cross))
simdb$sigma<-as.numeric(as.character(simdb$sigma))
simdb$nsv<-as.numeric(as.character(simdb$nsv))
#number of support vectors
simdb$trer<-as.numeric(as.character(simdb$trer))
simdb$crer<-as.numeric(as.character(simdb$crer))
simdb<-simdb[order(simdb$cost, simdb$cross),]

plot(simdb$trer, type="l")

simdb<-simdb[order(simdb$trer),]
svm.bm.mod<-get(paste("svm_",simdb[1,1], "_",
simdb[1,2], sep=""))

tabsim<-head(simdb, 10)
tabsim<-tabsim[, -c(3,6)]
rownames(tabsim)<-NULL
xtable(tabsim, digits=4)

tmp<-subset(simdb, simdb$cost==simdb$cost[1])
tmp<-tmp[order(tmp$cross),]

par(mar=c(5,4,4,5)+.1)
plot(tmp$cross, tmp$trer, type="l", col="red",
xlab="Cross Value", ylab="training error")
title( paste("Results with cost same as",
simdb$cost[1]))
par(new=TRUE)
plot(tmp$cross, tmp$crer,, type="l",
xaxt="n",yaxt="n",xlab="",ylab="")
axis(4)
mtext("Cross error",side=4,line=3)
legend("topleft",col=c("red","black"),lty=1,legend
=c("training","cross"), bg="white")

plot(tmp$cross, tmp$nsv, type="l", col="red",
xlab="Cross Value", ylab="training error")

simdb.1<-subset(simdb, simdb$cost>=1)
simdb.1<-subset(simdb.1, simdb.1$cost<101)
unique(simdb.1[,2])

plot(simdb.1$cost, simdb.1$cross)
#----------------------------
simdb.rgl<-simdb.1[,c(1,2,5)]
names(simdb.rgl)<-c("x","y","z")
simdb.rgl<-simdb.rgl[order(simdb.rgl$x,
simdb.rgl$y),]

simdb.1z<-
matrix(0,nrow=length(unique(simdb.rgl$x)),
ncol=length(unique(simdb.rgl$y)))

tmpx<-sort(unique(simdb.rgl$x))
tmpy<-sort(unique(simdb.rgl$y))

tmp<-1
for(i in 1:length(tmpx))
{
        for(j in 1:length(tmpx))
        {
                simdb.1z[i, j] <- simdb.rgl[tmp,3]
                tmp<-tmp+1
        }
}
simdb.1z

rgl.open()
material3d(col="white")
persp3d(x=unique(simdb.rgl$x),
y=unique(simdb.rgl$y), simdb.1z, col="lightblue",
xlab="Cost", ylab="Cross", zlab="Training error")
#play3d(spin3d(axis=c(0,0,1), rpm=10),
duration=30)
#---------------------------------------------

#---------------------------------------------
# Preparing best model

dim(aompromar06) #[1] 6031    1
dim(compdt)   #[1] 6031   15
names(compdt)

names(trainData)
svm.bm <- predict(svm.bm.mod, compdt[,-3])
table(svm.bm)

#install.packages("caret")
set.seed(1)

ptm.1<-proc.time()
rbfSVM<-train(x=trainData[,-c(3,15)],
y=trainData[,3],
        method="lssvmRadial",
        #preProc=c("center", "scale"),
        tuneLength = 8,
        trControl = trainControl(method =
"repeatedcv", repeats = 5),
        metric ="Kappa",
        fit=FALSE)
ptmf.2<-proc.time()-ptm.1

print(rbfSVM, printCall = FALSE)
densityplot(rbfSVM, metric ="Kappa", pch="|")
#densityplot(rbfSVM, metric ="ROC", pch="|")

svmPred <- predict(rbfSVM, testData[,-c(3,15)])
svmProbs <- predict(rbfSVM, testData[,-c(3,15)],
type = "prob")
confusionMatrix(svmPred, testData[,3])
plot(rbfSVM$resample$Accuracy,
rbfSVM$resample$Kappa)

#Variable Importance using caret
tab1<-varImp(rbfSVM, useModel=TRUE, scale=FALSE)
filterVarImp(x=trainData[,-c(3,15)],
y=trainData[,3])

tab1.1<-tab1[[1]]
getwd()
write.csv(tab1.1, "auc.csv", row.names=T)
sort(apply(tab1.1,1,mean))
plot(tab1)

#erasing "cld2, dist. and 06"
tab2<-tab1
rownames(tab2[[1]])<-gsub("cld2","dist.",
rownames(tab2[[1]]))
rownames(tab2[[1]])<-as.character(gsub("06","",
rownames(tab2[[1]])))
rownames(tab2[[1]])
tmp<-colnames(tab2[[1]])
colnames(tab2[[1]])<-letters[1:length(tmp)]
plot(tab2, cex=1)

#---------------------------------------------
load("unif_col.Rdata")
un_col<-tmp
un_col$code<-toupper(un_col$code)
un_col2<-un_col
un_col2$code[1:10]<-
c("a","b","c","d","e","f","A","B","E","F")

levels(testData$土地利用種)<-
c("a","b","c","d","e","f","A","B","E","F")

plot(aompromar06,

col=un_col[match(as.character(aompromar06@data
```

```
$土地利用種), rev(un_col$code)),"ncl"],
   bor=NA, axes=T)
title("Aomori UPA")

dev.new()
plot(aompromar06,
   col=un_col2[match(as.character(svm.bm),
rev(un_col2$code)),"ncl"],
   bor=NA, axes=T)
title("Aomori UPA, predicted data")

################################################
##
#plotting on google maps
install.packages("ReadImages")
library(RgoogleMaps)
library(PBSmapping)
library(maptools)

newaom<-aompromar06
head(newaom@data)

svm.bm.exp<-svm.bm
levels(svm.bm.exp)<-lev1

newaom@data$土地利用種<-as.character(svm.bm.exp)

plot(newaom,
   col=un_col[match(as.character(newaom@data$土地利
用種), rev(un_col2$code)),"ncl"],
   bor=NA, axes=T)

writePolyShape(newaom, "Aom_pred")
#############################################
#######

#--------------------------------------------
#Without long - lat
#Approach is not so good without this
information
trainData2<-trainData[,-c(1,2)]
testData2<-testData[,-c(1,2)]

svm.wll <- ksvm(myformula, data=trainData2,
kernel="rbfdot", kpar=list(sigma=0.01), C=1000,
cross=3)

svm.bm2 <- predict(svm.wll, compdt[,-c(1,2,3)])
tab.wll<-table(svm.bm2, compdt[,3])
apply(tab.wll,1,sum) /apply(tab.wll,2,sum)

dev.new()
plot(aompromar06,
   col=un_col2[match(as.character(svm.bm2),
rev(un_col2$code)),"ncl"],
   bor=NA, axes=T)
title("Aomori UPA2, predicted data")

#--------------------------------------------
#Using kpar automatic and increasing the cost,
the error is decreasing
getwd()
setwd("C:/Users/Kawai/Documents/carlos/PhD/SVM"
)
```

LENNARD-JONES CALCULATION FOR AOMORI MtA BETWEEN LAND USE AND SOCIO-ECONOMIC FACTORS

```
library(maptools)
library(sp)
library(xtable)
library(ggplot2)
library(gplots) #heatmap.2
library(spatstat)
library(rgdal) #sptransform

maindir<-"C:/Users/kuri73/Documents/carlos/PhD/experiments/last part"
setwd(maindir)

#-------------------------------------------------
# Loading
#-------------------------------------------------
load("info.Rdata")
#-------------------------------------------------
aomdb<-read.csv("aom_data.csv", header=T)
head(aomdb); dim(aomdb) #[1] 6031   16

options(encoding="UTF-8")
senddb<-read.csv("send_dbcomplete.csv", header=T)
head(senddb); dim(senddb) #[1] 30858    15

sapdb<-read.csv("sap_dbcomplete.csv", header=T)
head(sapdb); dim(sapdb) #[1] 72745    15

options(encoding="SHIFT-JIS")
save.image("info.Rdata")
#-------------------------------------------------
# Unifying
#-------------------------------------------------

names(aomdb)
aomdb$ID<-1:dim(aomdb)[1]

names(senddb)
names(sapdb)

nmdb<-c("ID", "long", "lat", "土地利用種",
  "cld2bus.st06", "cld2conv06", "cld2med.ins06",
"cld2oth.ins06",
  "cld2parks06", "cld2price06", "cld2pub.fl06", "
cld2spmk06", "cld2trst06", "cl.price", "pub.fl",
  "conv.nm")

aomdb <- aomdb[, match(nmdb, names(aomdb))]

#-------------------------------------------------
# Unifying names
#-------------------------------------------------
# Ordering

nmdb<-c("ID", "long", "lat", "土地利用種",
  "cld2bus.st06", "cld2conv06", "cld2med.ins06",
"cld2oth.ins06",
  "cld2parks06", "cld2price06", "cld2pub.fl06", "
cld2spmk06", "cld2trst06", "cl.price06", "pub.fl
",
  "conv.nm")

tmp<-match(nmdb, names(senddb))
senddb <- senddb[, tmp[!is.na(tmp)]]
head(senddb)

tmp<-match(nmdb, names(sapdb))
sapdb <- sapdb[, tmp[!is.na(tmp)]]
head(sapdb)
#-------------------------------------------------
#changing names

#Aomori

nmdb<-c("ID", "long", "lat", "土地利用種",
  "dbust", "dconv", "dmedins", "dothins",
  "dparks", "dprc", "dpubfl", "dspmk", "dtrst", "
prc", "pubfl",
  "convnm")

names(aomdb)<-nmdb

head(sapdb);names(sapdb);dim(sapdb)
head(senddb);names(senddb);dim(senddb)

nmdb<-c("ID", "long", "lat", "土地利用種",
  "dbust", "dconv", "dmedins",
  "dparks", "dprc", "dpubfl", "dspmk", "dtrst", "
prc", "pub.fl",
  "convnm")

names(sapdb)<-names(senddb)<-nmdb

head(sapdb)
head(senddb)
#-------------------------------------------------

#-------------------------------------------------
# Loading shapefiles

# Creating projection
llCRS <- CRS("+proj=longlat +ellps=WGS84")

aommaroi<-readShapePoly("C:/Users/kuri73/Documents/carlos/PhD/Aomori city/Aomma_roi.shp", proj4string=llCRS)
aomupa  <-readShapePoly("C:/Users/kuri73/Documents/carlos/PhD/image processing/aomori/06/aomori06prom.shp", proj4string=llCRS)
dim(aomupa) #[1] 6031    1

plot(aomupa)
unique(aomupa@data$土地利用種)
plot(aommaroi, col="green", add=T)
str(aommaroi)

aommaroi.tar<-sapply(slot(aommaroi, "polygons"),
function(x) sapply(slot(x, "Polygons"), slot, "area"))
aommaroi.tar1<-121

#-------------------------------------------------
# Starting Aomori
#-------------------------------------------------
head(aomdb);dim(aomdb) #[1] 6031   16
all(as.character(aomupa@data[,1]) %in% as.character(aomdb$土地利用種) ) #[1] TRUE

aomupa@data$ID<-1:dim(aomupa@data)[1]
head(aomupa@data)
coordinates(aomupa)

coords<-coordinates(aomupa)
pts <- SpatialPoints(coords, proj4string=CRS(proj4string(aomupa)))
plot(pts, add=T)

pts_polys <- over(pts, aommaroi)

plot(aomupa)
plot(aommaroi, col=rgb(0,1,0, alpha=0.5), add=T)
plot(pts[which(is.na(pts_polys)==FALSE)], add=T,
col="red")

aomin<-pts[which(is.na(pts_polys)==FALSE)]

coords<-SpatialPoints(coordinates(aomin),proj4string=llCRS)
aomin<-aomupa@data[which(is.na(pts_polys)==FALSE),]
names(aomin)[1]<-"LU"

maindb<- SpatialPointsDataFrame(coords, aomin)
writePointsShape(maindb,"aomccm.shp")

maindb@data<-aomdb[maindb@data$ID,]
aomdb.i<-maindb@data
dim(aomdb.i) #[1] 5261   16
```

```
dim(aomdb.i)[1]/dim(aomdb)[1] #[1] 0.8723263

#Interval
intval<-c(0, 1, 3, 5, 10, 20)

aomdb.i<-aomdb.i[,which(names(aomdb.i)!="dothins
")] #inside

tmpn<-c("dbust", "dconv", "dmedins", "dothins", "
dparks", "dprc", "dpubfl", "dspmk", "dtrst")
aomdb.icut<-aomdb.i

for(i in tmpn)
aomdb.icut[,i]<- cut(aomdb.i[,i], breaks=intval,
include.lowest=TRUE)
head(aomdb.icut)

for(i in tmpn)
levels(aomdb.icut[,i])<-c(1,2,3,4,5)
head(aomdb.icut)
levels(aomdb.icut$土地利用種)

#-------------------------------------------------
# To convert degrees to km
#-------------------------------------------------
eq1km<-0.0118685876079212 #calculated

setwd("C:/Users/kuri73/Documents/carlos/PhD/trans
_prob")

tmp<-dir()[grep("dist_", dir())]
tmp1<-gsub("dist_","",tmp)
tmp1<-gsub(".csv","",tmp1)

j<-1
for(i in tmp)
{
tmp2<-paste("d2",tmp1[j],sep="")
assign(tmp2, read.csv(i, head=TRUE) )
j=j+1
}
ls()
setwd(maindir)

#Converting into Kms
tmp<-ls()[grep("d2", ls())]
for (i in tmp)
{
tmp1<-get(i)
tmp1<-tmp1/eq1km
assign(i, tmp1)
}
ls()

#-----------------------------------
head(d2parks); dim(d2parks)
save.image("aom_chem.Rdata")
#-----------------------------------

#-------------------------------------------------
#Analyzing distances on different intervals
#-------------------------------------------------
tmp<-NULL
for(i in 1:nrow(d2parks))
{
 tmp<-rbind(tmp, table(cut(as.numeric(d2parks
[i,]), breaks = c(0,1,3,5,20))) )
}
cbind(tmp, apply(tmp, 1, sum))

head(cbind(tmp, apply(tmp, 1, sum)))

dim(tmp)
dim(unique(tmp))

heatmap(as.matrix(unique(tmp))) #takes time
head(data);dim(data)

#All the information
tmpob<-ls()[grep("d2",ls())]
for (i in tmpob)
{
        tmp<-NULL
        tmpf<-get(i)
        for(j in 1:nrow(tmpf))
        {
        tmp<-rbind(tmp, table(cut(as.numeric(tmpf
[j,]), breaks = c(0,1,3,5,20))) )
        }
assign(paste(i,".lv",sep=""), cbind(tmp, apply(tm
p, 1, sum)) )
}
ls()

tmp<-ls()[grep(".lv", ls())]
tmp<-tmp[-1]

#---------------------
head(d2bus.st.lv)
d2bus.st.lv
d2conv.lv
d2med.ins.lv
d2oth.ins.lv
d2parks.lv
d2price.lv
d2pub.fl.lv
d2spmk.lv
d2trst.lv
#---------------------

png(file="Aom_busst.png", width=800, height = 100
0, units = "px", pointsize = 16)
par(mfrow=c(4,1))
for(i in 1:4)
{
hist(d2bus.st.lv[,i], main="Aomori MtA UPA",
        xlab= paste("Number of bus stops in the in
terval",colnames(d2bus.st.lv)[i], "km"),
        freq=TRUE, col="gray", cex.axis=1.5, cex.l
ab=1.5, cex.main=1.5)
#curve(dnorm(x, mean=mean(d2bus.st.lv[,i]), sd=sd
(d2bus.st.lv[,i])), add=TRUE, col="darkblue", lwd
=2)
#lines(density(d2bus.st.lv[,i]))
}
dev.off()

getwd()
png(file="Aom_busst.png", width=800, height = 100
0, units = "px", pointsize = 16)
tmpdb<-d2bus.st.lv
par(mfrow=c(4,1))
for(i in 1:4)
{
hist(tmpdb[,i], main="Aomori MtA UPA",
        xlab= paste("Number of bus stops in the in
terval",colnames(tmpdb)[i], "km"),
        freq=TRUE, col="gray", cex.axis=1.5, cex.l
ab=1.5, cex.main=1.5)
}
dev.off()
head(tmpdb)

#calculating expected value
exval<-function(x)
{
tmpu<-t(as.matrix(sort(unique(x))))
tmpp<-as.matrix(as.numeric(prop.table(table(x))))
return(tmpu %*% tmpp)
}

tmpt<-NULL
for(i in tmp)
{
cat(i, "\n")
print(apply(get(i)[,c(1,2,3,4)], 2, mean))
cat("-------------------\n")
#tmpt<-rbind(tmpt, apply(get(i)[,c(1,2,3,4)], 2,
mean) )
tmpt<-rbind(tmpt, apply(get(i)[,c(1,2,3,4)], 2, e
```

```
xval) )
}

rownames(tmpt)<-tmp
xtable(tmpt, digis=2)

apply(d2bus.st.lv[,c(1,2,3,4)], 2, exval)


#------------------------------------------------
# Loading AUC
#------------------------------------------------
AUC<-read.csv("C:/Users/kuri73/Documents/carlos/P
hD/experiments/third part/auc.csv", header=T, ro
w.names=1)
rownames(AUC) <- c("long", "lat", "dbust", "dconv
", "dmedins", "dothins",
  "dparks", "dprc", "dpubfl", "dspmk", "dtrst", "
prc", "pub.fl")

#------------------------------------------------
# Starting extraction
#------------------------------------------------
head(aomdb.icut[,c(5:13)]);dim(aomdb.icut[,c(5:1
3)]) #[1] 5261    9
dim(unique(aomdb.icut[,c(5:13)])) #[1] 56  9


# type a or 1
aomdb.lua<-subset(aomdb.icut, aomdb.icut$土地利用種
==1)
head(aomdb.lua)

AUC.a<-as.data.frame(cbind(type=rownames(AUC), AU
C=AUC$a))
AUC.a$AUC<-as.numeric(as.character(AUC.a$AUC))
AUC.a<-AUC.a[order(AUC.a$AUC),]

par(mar=c(6.1,4.1,3.1,2.1)) #b,l,u,r
barplot(AUC.a$AUC, names.arg=AUC.a$type, ylim=c
(0,1), las=2)

head(aomdb.lua)


# type 7 or 7
aomdb.lubs<-subset(aomdb.icut, aomdb.icut$土地利用
種==7)
head(aomdb.lubs)

AUC.e<-as.data.frame(cbind(type=rownames(AUC), AU
C=AUC$e))
AUC.e$AUC<-as.numeric(as.character(AUC.e$AUC))
AUC.e<-AUC.e[order(AUC.e$AUC),]

labs<-c("prices", "Public fl.", "Distance to park
s", "Distance to train st.", "Distance to bus st.
",
       "Distance to supermarket", "Distance to po
l. price", "Distance to oth. institutions",
       "Distance to public fl.", "Distance to me
d. institutions", "Longitude", "Distance to conv.
st", "Latitude")

AUC.e$type <- factor(AUC.e$type, levels=unique(a
s.character(AUC.e$type)) )
labs<-
c<-ggplot(AUC.e, aes(type, AUC)) + labs(title = "
AUC for residential area", x = "Socio-economic fa
ctors", y = "AUC") +
   theme(text = element_text(size=18))
c+geom_bar(labels=labs) + scale_x_discrete(labels
=labs) + coord_flip()

ls()

head(aomdb.icut); dim(aomdb.icut) #[1] 5261   16
xtable(prop.table(table(aomdb.icut$土地利用種))*10
0, digits=3)
#           1            2            5
6           7            9
#0.0505607299 0.0140657670 0.0178673256 0.0058924
159 0.6768675157 0.0543622885
#           A            B            E
F
#0.1338148641 0.0190077932 0.0003801559 0.0271811
443

tmp<-aomdb.lubs[,c(5:13)]
dim(tmp) #[1] 3561    9
dim(tmp)[1]/dim(aomdb)[1] #[1] 0.5904493

tmp1<-unique(data.matrix(unique(tmp)))
dim(tmp1) #[1] 40  9

dim(unique(aomdb.icut[,c(5:13)])) #[1] 56  9
dim(tmp1)[1]/dim(unique(aomdb.icut[,c(5:13)]))[1]
#[1] 0.7142857

ls()

#ordering according AUC
tmph<-match(AUC.e$type, colnames(tmp1))
tmph<-tmph[!is.na(tmph)]
tmp1<-tmp1[,tmph]
head(tmp1);dim(tmp1)
colnames(tmp1)

tabpt<-cbind(aucnm=as.character(AUC.e$type), lab
s)
heatmap.2(tmp1, trace="none", col=greenred(10)) #
takes time

#table(tmp1)/sum(table(tmp1))
#         1          2          3
#0.73333333 0.23055556 0.03611111

labs[tmph]
#------------------------------------------------
#Starting molecule, using residential area
#------------------------------------------------
maindir.1<-"C:/Users/kuri73/Documents/carlos/PhD/
popul_detail/detailed info"
setwd(maindir.1)
dir()

bus.st<-readShapePoints(paste(maindir.1, "/bus_st
op/P11-10_02-jgd_ED01.shp",sep=""), proj4string=l
lCRS) #convinience stores
conv<-readShapePoints(paste(maindir.1, "/convenie
nce store/aom_convst.shp",sep=""), proj4string=ll
CRS) #convinience stores
med.ins<-readShapePoints(paste(maindir.1, "/medic
al instit/P04-10_02-g_MedicalInstitution.shp",sep
=""), proj4string=llCRS)
oth.ins<-readShapePoints(paste(maindir.1, "/other
s/L01-12_02_DA01.shp",sep=""), proj4string=llCRS)
#other institutions
parks<-readShapePoints(paste(maindir.1, "/parks/P
13-11_02.shp",sep=""), proj4string=llCRS)
price<-readShapePoints(paste(maindir.1, "/price/L
01-06_02_DA01.shp",sep=""), proj4string=llCRS)
pub.fl<-readShapePoints(paste(maindir.1, "/public
facilities/P02-06_02-g_PublicFacility.shp",sep="
"), proj4string=llCRS) #public facilities
spmk<-readShapePoints(paste(maindir.1, "/supermar
kets/aomori.shp",sep=""), proj4string=llCRS) #sup
ermarkets
trst<-readShapePoints(paste(maindir.1, "/stations
/jstations.shp",sep=""), proj4string=llCRS) #trai
n stations

plot(aomupa, border="gray")
plot(aommaroi, add=T, col=rgb(1,0,0, alpha=0.5))
#plot(spmk, add=T, col="green", pch=19)
plot(bus.st, add=T, col="green", pch=19)
```

```
#-------------------------------------------------
# Calculating combinations
#-------------------------------------------------
sef<-c("bus.st","conv","med.ins","oth.ins","parks
","price","pub.fl","spmk","trst")

elemn<-c("lu", sef)
celem<-combn(elemn, 2)
celem #45
#-------------------------------------------------

#-------------------------------------------------
#Bus
#-------------------------------------------------
pts <- SpatialPoints(coordinates(bus.st), proj4st
ring=CRS(proj4string(aomupa)))
pts_polys <- over(bus.st, aommaroi)
aom.cbs<-pts[which(is.na(pts_polys)==FALSE)]

#-------------------------------------------------
dblubs<-cbind(aomdb.lubs[,c("long", "lat")], type
= "e")
head(dblubs)
tmp<-cbind(coordinates(aom.cbs),type="bus")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dblubs<-rbind(dblubs, tmp)
dblubs[,1]<-as.numeric(dblubs[,1])
dblubs[,2]<-as.numeric(dblubs[,2])
head(dblubs)
dim(dblubs)
dim(unique(dblubs))

tmpc<-rep("green", dim(dblubs)[1])
tmpc[which(dblubs$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dblubs$long, dblubs$lat, col=tmpc, pch=20,
main="Core and bus stops", xlab="Longitude", ylab
="latitude")

tmp<-as.data.frame(dblubs[,3])

coords<-SpatialPoints(dblubs[,c("long","lat")],pr
oj4string=llCRS)
aomlubs.e<- SpatialPointsDataFrame(coords, tmp) #
export

aomlubs.e<- spTransform(aomlubs.e, CRS("+proj=utm
+zone=54 ellps=WGS84")) #converting into mt
plot(aomlubs.e, col=tmpc)

aomlubs.p<-as(aomlubs.e, "ppp")
class(aomlubs.p) #[1] "ppp"
aomlubs.p<-unique.ppp(aomlubs.p)

plot(split(aomlubs.p), pch=20)

rhat <- min(nndist(aomlubs.p)) #nearest neighbor
rhat <- rhat * 0.99999
#rhat<-1
#ppm(aomlubs.p, ~1, Strauss(r = rhat), correction
="translate")
#Relevant coefficients:
#Interaction
#  -15.87649

aomebs.lj<-ppm(aomlubs.p, ~1, LennardJones(), cor
rection="translate", gcontrol=list(maxit=1e4), rb
ond=rhat)
coef(summary(aomebs.lj))
str(aomebs.lj)
s<-aomebs.lj$interaction$par$sigma0
t<-aomebs.lj$internal$glmfit$control$epsilon
aomebs.st<-ppm(aomlubs.p, ~1, Strauss(r = rhat),
correction="translate")
plot(density(aomlubs.p,150))
plot(Kest(aomlubs.p))
E<-envelope(aomlubs.p, Kest, nsim=40)
plot(E)
#-------------------------------------------------

#-------------------------------------------------
#conv
#-------------------------------------------------
pts <- SpatialPoints(coordinates(conv), proj4stri
ng=CRS(proj4string(aomupa)))
pts_polys <- over(conv, aommaroi)
aom.cconv<-pts[which(is.na(pts_polys)==FALSE)]

#-------------------------------------------------
dbluconv<-cbind(aomdb.lubs[,c("long", "lat")], ty
pe= "e")
head(dbluconv);dim(dbluconv)
tmp<-cbind(coordinates(aom.cconv),type="conv")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dbluconv<-rbind(dbluconv, tmp)
dbluconv[,1]<-as.numeric(dbluconv[,1])
dbluconv[,2]<-as.numeric(dbluconv[,2])
head(dbluconv);dim(dbluconv)
dim(unique(dbluconv))

#----------
tmpc<-rep("green", dim(dbluconv)[1])
tmpc[which(dbluconv$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dbluconv$long, dbluconv$lat, col=tmpc, pch
=20, main="Core and bus stops", xlab="Longitude",
ylab="latitude")
#----------

tmp<-as.data.frame(dbluconv[,3])
coords<-SpatialPoints(dbluconv[,c("long","lat")],
proj4string=llCRS)
aomluconv.e<- SpatialPointsDataFrame(coords, tmp)
#export

aomluconv.e<- spTransform(aomluconv.e, CRS("+proj
=utm +zone=54 ellps=WGS84")) #converting into mt
plot(aomluconv.e, col=tmpc, pch=20)

aomlubconv.p<-as(aomluconv.e, "ppp")
class(aomlubconv.p) #[1] "ppp"
aomlubconv.p<-unique.ppp(aomlubconv.p)

plot(split(aomlubconv.p), pch=20)

rhat <- min(nndist(aomlubconv.p)) #nearest neighb
or
rhat <- rhat * 0.99999

aomluconv.lj<-ppm(aomlubconv.p, ~1, LennardJones
(), correction="translate", gcontrol=list(maxit=1
e3))
coef(summary(aomluconv.lj))
str(aomluconv.lj)
s<-aomluconv.lj$interaction$par$sigma0
t<-aomluconv.lj$internal$glmfit$control$epsilon
aomluconv.st<-ppm(aomlubconv.p, ~1, Strauss(r = r
hat), correction="translate")
plot(density(aomlubconv.p,150))
plot(Kest(aomlubconv.p))
E<-envelope(aomlubconv.p, Kest, nsim=40)
plot(E)
#-------------------------------------------------

#-------------------------------------------------
celem[,3]
#medins
#-------------------------------------------------
pts <- SpatialPoints(coordinates(med.ins), proj4s
tring=CRS(proj4string(aomupa)))
pts_polys <- over(med.ins, aommaroi)
aom.cmed.ins<-pts[which(is.na(pts_polys)==FALSE)]

#-------------------------------------------------
dblumed.ins<-cbind(aomdb.lubs[,c("long", "lat")],
type= "e")
```

```
head(dblumed.ins);dim(dblumed.ins)
tmp<-cbind(coordinates(aom.cmed.ins),type="med.in
s")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dblumed.ins<-rbind(dblumed.ins, tmp)
dblumed.ins[,1]<-as.numeric(dblumed.ins[,1])
dblumed.ins[,2]<-as.numeric(dblumed.ins[,2])
head(dblumed.ins);dim(dblumed.ins)
dim(unique(dblumed.ins))

#----------
tmpc<-rep("green", dim(dblumed.ins)[1])
tmpc[which(dblumed.ins$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dblumed.ins$long, dblumed.ins$lat, col=tmp
c, pch=20, main="Core and bus stops", xlab="Longi
tude", ylab="latitude")
#----------

tmp<-as.data.frame(dblumed.ins[,3])
coords<-SpatialPoints(dblumed.ins[,c("long","lat
")],proj4string=llCRS)
aomlumed.ins.e<- SpatialPointsDataFrame(coords, t
mp) #export

aomlumed.ins.e<- spTransform(aomlumed.ins.e, CRS
("+proj=utm +zone=54 ellps=WGS84")) #med.insertin
g into mt
plot(aomlumed.ins.e, col=tmpc, pch=20)

aomlubmed.ins.p<-as(aomlumed.ins.e, "ppp")
class(aomlubmed.ins.p) #[1] "ppp"
aomlubmed.ins.p<-unique.ppp(aomlubmed.ins.p)

plot(split(aomlubmed.ins.p), pch=20)

rhat <- min(nndist(aomlubmed.ins.p)) #nearest nei
ghbor
rhat <- rhat * 0.99999

aomlumed.ins.lj<-ppm(aomlubmed.ins.p, ~1, Lennard
Jones(), correction="translate", gcontrol=list(ma
xit=1e3))
#coef(summary(aomlumed.ins.lj))
#str(aomlumed.ins.lj)
s<-aomlumed.ins.lj$interaction$par$sigma0
t<-aomlumed.ins.lj$internal$glmfit$control$epsilo
n
aomlumed.ins.st<-ppm(aomlubmed.ins.p, ~1, Strauss
(r = rhat), correction="translate")
plot(density(aomlubmed.ins.p,150))
#plot(Kest(aomlubmed.ins.p))
E<-envelope(aomlubmed.ins.p, Kest, nsim=40)
plot(E)
#-------------------------------------------------

#-------------------------------------------------
celem[,4]
#oth.ins
#-------------------------------------------------
pts <- SpatialPoints(coordinates(oth.ins), proj4s
tring=CRS(proj4string(aomupa)))
pts_polys <- over(oth.ins, aommaroi)
aom.coth.ins<-pts[which(is.na(pts_polys)==FALSE)]

#-------------------------------------------------
dbluoth.ins<-cbind(aomdb.lubs[,c("long", "lat")],
type= "e")
head(dbluoth.ins);dim(dbluoth.ins)
tmp<-cbind(coordinates(aom.coth.ins),type="oth.in
s")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dbluoth.ins<-rbind(dbluoth.ins, tmp)
dbluoth.ins[,1]<-as.numeric(dbluoth.ins[,1])
dbluoth.ins[,2]<-as.numeric(dbluoth.ins[,2])
head(dbluoth.ins);dim(dbluoth.ins)
dim(unique(dbluoth.ins))

#----------
tmpc<-rep("green", dim(dbluoth.ins)[1])
tmpc[which(dbluoth.ins$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dbluoth.ins$long, dbluoth.ins$lat, col=tmp
c, pch=20, main="Core and bus stops", xlab="Longi
tude", ylab="latitude")
#----------

tmp<-as.data.frame(dbluoth.ins[,3])
coords<-SpatialPoints(dbluoth.ins[,c("long","lat
")],proj4string=llCRS)
aomluoth.ins.e<- SpatialPointsDataFrame(coords, t
mp) #export

aomluoth.ins.e<- spTransform(aomluoth.ins.e, CRS
("+proj=utm +zone=54 ellps=WGS84")) #oth.insertin
g into mt
#plot(aomluoth.ins.e, col=tmpc, pch=20)

aomluboth.ins.p<-as(aomluoth.ins.e, "ppp")
#class(aomluboth.ins.p) #[1] "ppp"
aomluboth.ins.p<-unique.ppp(aomluboth.ins.p)

plot(split(aomluboth.ins.p), pch=20)

rhat <- min(nndist(aomluboth.ins.p)) #nearest nei
ghbor
rhat <- rhat * 0.99999

aomluoth.ins.lj<-ppm(aomluboth.ins.p, ~1, Lennard
Jones(), correction="translate", gcontrol=list(ma
xit=1e3))
#coef(summary(aomluoth.ins.lj))
#str(aomluoth.ins.lj)
#s<-aomluoth.ins.lj$interaction$par$sigma0
#t<-aomluoth.ins.lj$internal$glmfit$control$epsil
on
aomluoth.ins.st<-ppm(aomluboth.ins.p, ~1, Strauss
(r = rhat), correction="translate")
#plot(density(aomluboth.ins.p,150))
#plot(Kest(aomluboth.ins.p))
E<-envelope(aomluboth.ins.p, Kest, nsim=40)
plot(E)
#-------------------------------------------------

#-------------------------------------------------
celem[,5]
#parks
#-------------------------------------------------
pts <- SpatialPoints(coordinates(parks), proj4str
ing=CRS(proj4string(aomupa)))
pts_polys <- over(parks, aommaroi)
aom.cparks<-pts[which(is.na(pts_polys)==FALSE)]

#-------------------------------------------------
dbluparks<-cbind(aomdb.lubs[,c("long", "lat")], t
ype= "e")
head(dbluparks);dim(dbluparks)
tmp<-cbind(coordinates(aom.cparks),type="parks")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dbluparks<-rbind(dbluparks, tmp)
dbluparks[,1]<-as.numeric(dbluparks[,1])
dbluparks[,2]<-as.numeric(dbluparks[,2])
head(dbluparks);dim(dbluparks)
dim(unique(dbluparks))

#----------
tmpc<-rep("green", dim(dbluparks)[1])
tmpc[which(dbluparks$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dbluparks$long, dbluparks$lat, col=tmpc, p
ch=20, main="Core and bus stops", xlab="Longitude
", ylab="latitude")
#----------
```

```
tmp<-as.data.frame(dbluparks[,3])
coords<-SpatialPoints(dbluparks[,c("long","lat
")],proj4string=llCRS)
aomluparks.e<- SpatialPointsDataFrame(coords, tm
p) #export

aomluparks.e<- spTransform(aomluparks.e, CRS("+pr
oj=utm +zone=54 ellps=WGS84")) #parkserting into
mt
#plot(aomluparks.e, col=tmpc, pch=20)

aomlubparks.p<-as(aomluparks.e, "ppp")
#class(aomlubparks.p) #[1] "ppp"
aomlubparks.p<-unique.ppp(aomlubparks.p)

plot(split(aomlubparks.p), pch=20)

rhat <- min(nndist(aomlubparks.p)) #nearest neigh
bor
rhat <- rhat * 0.99999

aomluparks.lj<-ppm(aomlubparks.p, ~1, LennardJone
s(), correction="translate", gcontrol=list(maxit=
1e3))
#coef(summary(aomluparks.lj))
#str(aomluparks.lj)
#s<-aomluparks.lj$interaction$par$sigma0
#t<-aomluparks.lj$internal$glmfit$control$epsilon
aomluparks.st<-ppm(aomlubparks.p, ~1, Strauss(r =
rhat), correction="translate")
#plot(density(aomlubparks.p,150))
#plot(Kest(aomlubparks.p))
E<-envelope(aomlubparks.p, Kest, nsim=40)
plot(E)
#-----------------------------------------------

#-----------------------------------------------
celem[,6]
#price
#-----------------------------------------------
pts <- SpatialPoints(coordinates(price), proj4str
ing=CRS(proj4string(aomupa)))
pts_polys <- over(price, aommaroi)
aom.cprice<-pts[which(is.na(pts_polys)==FALSE)]

#-----------------------------------------------
dbluprice<-cbind(aomdb.lubs[,c("long", "lat")], t
ype= "e")
head(dbluprice);dim(dbluprice)
tmp<-cbind(coordinates(aom.cprice),type="price")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dbluprice<-rbind(dbluprice, tmp)
dbluprice[,1]<-as.numeric(dbluprice[,1])
dbluprice[,2]<-as.numeric(dbluprice[,2])
head(dbluprice);dim(dbluprice)
dim(unique(dbluprice))

#----------
tmpc<-rep("green", dim(dbluprice)[1])
tmpc[which(dbluprice$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dbluprice$long, dbluprice$lat, col=tmpc, p
ch=20, main="Core and bus stops", xlab="Longitude
", ylab="latitude")
#----------

tmp<-as.data.frame(dbluprice[,3])
coords<-SpatialPoints(dbluprice[,c("long","lat
")],proj4string=llCRS)
aomluprice.e<- SpatialPointsDataFrame(coords, tm
p) #export

aomluprice.e<- spTransform(aomluprice.e, CRS("+pr
oj=utm +zone=54 ellps=WGS84")) #priceerting into
mt
#plot(aomluprice.e, col=tmpc, pch=20)

aomlubprice.p<-as(aomluprice.e, "ppp")
#class(aomlubprice.p) #[1] "ppp"
aomlubprice.p<-unique.ppp(aomlubprice.p)

plot(split(aomlubprice.p), pch=20)

rhat <- min(nndist(aomlubprice.p)) #nearest neigh
bor
rhat <- rhat * 0.99999

aomluprice.lj<-ppm(aomlubprice.p, ~1, LennardJone
s(), correction="translate", gcontrol=list(maxit=
1e3))
#coef(summary(aomluprice.lj))
#str(aomluprice.lj)
#s<-aomluprice.lj$interaction$par$sigma0
#t<-aomluprice.lj$internal$glmfit$control$epsilon
aomluprice.st<-ppm(aomlubprice.p, ~1, Strauss(r =
rhat), correction="translate")
#plot(density(aomlubprice.p,150))
#plot(Kest(aomlubprice.p))
E<-envelope(aomlubprice.p, Kest, nsim=40)
plot(E)
#-----------------------------------------------

#-----------------------------------------------
celem[,7]
#pub.fl
#-----------------------------------------------
pts <- SpatialPoints(coordinates(pub.fl), proj4st
ring=CRS(proj4string(aomupa)))
pts_polys <- over(pub.fl, aommaroi)
aom.cpub.fl<-pts[which(is.na(pts_polys)==FALSE)]

#-----------------------------------------------
dblupub.fl<-cbind(aomdb.lubs[,c("long", "lat")],
type= "e")
head(dblupub.fl);dim(dblupub.fl)
tmp<-cbind(coordinates(aom.cpub.fl),type="pub.fl
")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dblupub.fl<-rbind(dblupub.fl, tmp)
dblupub.fl[,1]<-as.numeric(dblupub.fl[,1])
dblupub.fl[,2]<-as.numeric(dblupub.fl[,2])
head(dblupub.fl);dim(dblupub.fl)
dim(unique(dblupub.fl))
dblupub.fl<-unique(dblupub.fl)

#----------
tmpc<-rep("green", dim(dblupub.fl)[1])
tmpc[which(dblupub.fl$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dblupub.fl$long, dblupub.fl$lat, col=tmpc,
pch=20, main="Core and bus stops", xlab="Longitud
e", ylab="latitude")
#----------

tmp<-as.data.frame(dblupub.fl[,3])
coords<-SpatialPoints(dblupub.fl[,c("long","lat
")],proj4string=llCRS)
aomlupub.fl.e<- SpatialPointsDataFrame(coords, tm
p) #export

aomlupub.fl.e<- spTransform(aomlupub.fl.e, CRS("+
proj=utm +zone=54 ellps=WGS84")) #pub.flerting in
to mt
#plot(aomlupub.fl.e, col=tmpc, pch=20)

aomlubpub.fl.p<-as(aomlupub.fl.e, "ppp")
#class(aomlubpub.fl.p) #[1] "ppp"
aomlubpub.fl.p<-unique.ppp(aomlubpub.fl.p)

plot(split(aomlubpub.fl.p), pch=20)

rhat <- min(nndist(aomlubpub.fl.p)) #nearest neig
hbor
rhat <- rhat * 0.99999

aomlupub.fl.lj<-ppm(aomlubpub.fl.p, ~1, LennardJo
```

```
nes(), correction="translate", gcontrol=list(maxi
t=1e3))
#coef(summary(aomlupub.fl.lj))
#str(aomlupub.fl.lj)
#s<-aomlupub.fl.lj$interaction$par$sigma0
#t<-aomlupub.fl.lj$internal$glmfit$control$epsilo
n
aomlupub.fl.st<-ppm(aomlubpub.fl.p, ~1, Strauss(r
= rhat), correction="translate")
#plot(density(aomlubpub.fl.p,150))
#plot(Kest(aomlubpub.fl.p))
E<-envelope(aomlubpub.fl.p, Kest, nsim=40)
plot(E)
#------------------------------------------------

#------------------------------------------------
celem[,8]
#spmk
#------------------------------------------------
pts <- SpatialPoints(coordinates(spmk), proj4stri
ng=CRS(proj4string(aomupa)))
pts_polys <- over(spmk, aommaroi)
aom.cspmk<-pts[which(is.na(pts_polys)==FALSE)]

#------------------------------------------------
dbluspmk<-cbind(aomdb.lubs[,c("long", "lat")], ty
pe= "e")
head(dbluspmk);dim(dbluspmk)
tmp<-cbind(coordinates(aom.cspmk),type="spmk")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dbluspmk<-rbind(dbluspmk, tmp)
dbluspmk[,1]<-as.numeric(dbluspmk[,1])
dbluspmk[,2]<-as.numeric(dbluspmk[,2])
head(dbluspmk);dim(dbluspmk)
dim(unique(dbluspmk))
dbluspmk<-unique(dbluspmk)

#----------
tmpc<-rep("green", dim(dbluspmk)[1])
tmpc[which(dbluspmk$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dbluspmk$long, dbluspmk$lat, col=tmpc, pch
=20, main="Core and bus stops", xlab="Longitude",
ylab="latitude")
#----------

tmp<-as.data.frame(dbluspmk[,3])
coords<-SpatialPoints(dbluspmk[,c("long","lat")],
proj4string=llCRS)
aomluspmk.e<- SpatialPointsDataFrame(coords, tmp)
#export

aomluspmk.e<- spTransform(aomluspmk.e, CRS("+proj
=utm +zone=54 ellps=WGS84")) #spmkerting into mt
#plot(aomluspmk.e, col=tmpc, pch=20)

aomlubspmk.p<-as(aomluspmk.e, "ppp")
#class(aomlubspmk.p) #[1] "ppp"
aomlubspmk.p<-unique.ppp(aomlubspmk.p)

plot(split(aomlubspmk.p), pch=20)

rhat <- min(nndist(aomlubspmk.p)) #nearest neighb
or
rhat <- rhat * 0.99999

aomluspmk.lj<-ppm(aomlubspmk.p, ~1, LennardJones
(), correction="translate", gcontrol=list(maxit=1
e3))
#coef(summary(aomluspmk.lj))
#str(aomluspmk.lj)
#s<-aomluspmk.lj$interaction$par$sigma0
#t<-aomluspmk.lj$internal$glmfit$control$epsilon
aomluspmk.st<-ppm(aomlubspmk.p, ~1, Strauss(r = r
hat), correction="translate")
#plot(density(aomlubspmk.p,150))
#plot(Kest(aomlubspmk.p))
#E<-envelope(aomlubspmk.p, Kest, nsim=40)
```

```
#plot(E)
#------------------------------------------------

#------------------------------------------------
celem[,9]
#trst
#------------------------------------------------
pts <- SpatialPoints(coordinates(trst), proj4stri
ng=CRS(proj4string(aomupa)))
pts_polys <- over(trst, aommaroi)
aom.ctrst<-pts[which(is.na(pts_polys)==FALSE)]

#------------------------------------------------
dblutrst<-cbind(aomdb.lubs[,c("long", "lat")], ty
pe= "e")
head(dblutrst);dim(dblutrst)
tmp<-cbind(coordinates(aom.ctrst),type="trst")
tmp<-as.data.frame(tmp)
names(tmp)[c(1,2)]<-c("long", "lat")
dblutrst<-rbind(dblutrst, tmp)
dblutrst[,1]<-as.numeric(dblutrst[,1])
dblutrst[,2]<-as.numeric(dblutrst[,2])
head(dblutrst);dim(dblutrst)
dim(unique(dblutrst))
dblutrst<-unique(dblutrst)

#----------
tmpc<-rep("green", dim(dblutrst)[1])
tmpc[which(dblutrst$type=="e")]<-"red"
plot(aomupa, border="gray", axes=T)
title("Aomori MtA UPA core (Land use type e)")
points(dblutrst$long, dblutrst$lat, col=tmpc, pch
=20, main="Core and bus stops", xlab="Longitude",
ylab="latitude")
#----------

tmp<-as.data.frame(dblutrst[,3])
coords<-SpatialPoints(dblutrst[,c("long","lat")],
proj4string=llCRS)
aomlutrst.e<- SpatialPointsDataFrame(coords, tmp)
#export

aomlutrst.e<- spTransform(aomlutrst.e, CRS("+proj
=utm +zone=54 ellps=WGS84")) #trsterting into mt
#plot(aomlutrst.e, col=tmpc, pch=20)

aomlubtrst.p<-as(aomlutrst.e, "ppp")
#class(aomlubtrst.p) #[1] "ppp"
aomlubtrst.p<-unique.ppp(aomlubtrst.p)

plot(split(aomlubtrst.p), pch=20)

rhat <- min(nndist(aomlubtrst.p)) #nearest neighb
or
rhat <- rhat * 0.99999

aomlutrst.lj<-ppm(aomlubtrst.p, ~1, LennardJones
(), correction="translate", gcontrol=list(maxit=1
e3))
#coef(summary(aomlutrst.lj))
#str(aomlutrst.lj)
#s<-aomlutrst.lj$interaction$par$sigma0
#t<-aomlutrst.lj$internal$glmfit$control$epsilon
aomlutrst.st<-ppm(aomlubtrst.p, ~1, Strauss(r = r
hat), correction="translate")
#plot(density(aomlubtrst.p,150))
#plot(Kest(aomlubtrst.p))
#E<-envelope(aomlubtrst.p, Kest, nsim=40)
#plot(E)
#------------------------------------------------

#------------------------------------------------
# interlude
#------------------------------------------------
dblubus.st<-dblubs # unifying .....
tmpdbn<-ls()[grep("dblu",ls())]
tmpdbn<-tmpdbn[-which(tmpdbn=="dblubs")]
#[1] "dblubus.st"  "dbluconv"    "dblumed.ins" "d
bluoth.ins" "dbluparks"
#[6] "dbluprice"   "dblupub.fl"  "dbluspmk"    "d
```

```
blutrst"

sef
#[1] "bus.st"  "conv"    "med.ins" "oth.ins" "par
ks"   "price"   "pub.fl"
#[8] "spmk"    "trst"
#-----------------------------------------------

#-----------------------------------------------
celem[,10]
#[1] "oth.ins" "price"
#-----------------------------------------------
for (i in 33:dim(celem)[2])
{
#       i=10
        tmpcl1<- celem[1,i]
        tmpcl2<- celem[2,i]

        tmp<-get(paste("dblu", tmpcl1, sep=""))
        tmpcl1<-tmp[-which(tmp$type=="e"),]
        tmp<-get(paste("dblu", tmpcl2, sep=""))
        tmpcl2<-tmp[-which(tmp$type=="e"),]

#-----------------------------------------------
        tmp<-rbind(tmpcl1, tmpcl2)
        tmp[,3]<-as.character(tmp[,3])
        head(tmp); tail(tmp)
        assign(paste("db", celem[1,i],".",celem[2,
i], sep=""), tmp)

        names(tmp)[c(1,2)]<-c("long", "lat")
        tmp[,1]<-as.numeric(as.character(tmp[,1]))
        tmp[,2]<-as.numeric(as.character(tmp[,2]))
        tmp[,3]
        dim(tmp)

        tmp<-unique(tmp); dim(tmp)
        tmp3<-as.data.frame(tmp[,3])
        coords<-SpatialPoints(tmp[,c("long","lat
")],proj4string=llCRS)

        tmp.e<- SpatialPointsDataFrame(coords, tmp
3) #export
        tmp.e<- spTransform(tmp.e, CRS("+proj=utm
+zone=54 ellps=WGS84")) #trsterting into mt
        assign(paste("db", celem[1,i],".",celem[2,
i],".e", sep=""), tmp.e)

        tmp.p<-as(tmp.e, "ppp")
        tmp.p<-unique.ppp(tmp.p)
        assign(paste("db", celem[1,i],".",celem[2,
i],".p", sep=""), tmp.p)

        setwd("C:/Users/kuri73/Desktop/New folder
")
        png(paste("db", celem[1,i],".",celem[2,i],
".png", sep=""), width = 1280, height = 982, unit
s = "px")
        plot(split(tmp.p), pch=20)
        dev.off()

        setwd(maindir)
        rhat <- min(nndist(tmp.p)) #nearest neighb
or
        rhat <- rhat * 0.99999
        rhat <- ifelse(rhat==0, 0.99999, rhat)

        unique.ppp(tmp.p)

        tmp.lj<-ppm(tmp.p, ~1, LennardJones(), cor
rection="translate", gcontrol=list(maxit=1e3))
        str(tmp.lj)
        summary(tmp.lj)
        assign(paste("db", celem[1,i],".",celem[2,
i],".lj", sep=""), tmp.lj)
        tmp.st<-ppm(tmp.p, ~1, Strauss(r = rhat),
correction="translate", gcontrol=list(maxit=1e3))
        assign(paste("db", celem[1,i],".",celem[2,
i],".st", sep=""), tmp.st)
```

```
        save.image("info.Rdata")
}

#-----------------------------------------------

tmplj<-ls()[grep(".lj$", ls())]
tmplj<-tmplj[-which(tmplj=="tmp.lj")]

tmpst<-paste(gsub(".lj", "", tmplj), ".st", sep="
")


dbres.st<-NULL
for(i in tmpst)
{
cat(i)
        tmp<-get(i)
        tmps<-summary(tmp)
        dbres.st<-rbind(dbres.st, c(i, as.numeric
(coefficients(tmp)), tmps$data$intensity))
}
dbres.st<-as.data.frame(dbres.st, stringsAsFactor
s = FALSE)
names(dbres.st)<-c("name", "intercept", "interact
ion", "intensity")
dbres.st

dbres.lj<-NULL
for(i in tmplj)
{
#i="dbbus.st.conv.lj"
        tmp<-get(i)
        coef(summary(tmp))
        tmps<-summary(tmp)
        str(tmps)

dbres.lj<-rbind(dbres.lj, c(i, coefficients(tmps)
[,1], #intercept interact.1 interact.2
        as.numeric(tmps$trend$value), #  beta_spmk
beta_trst
        as.numeric(tmps$interaction$sensible$print
able)) ) #   sigma  epsilon
cat(i, "\n")
}
dbres.lj<-as.data.frame(dbres.lj, stringsAsFactor
s = FALSE)
names(dbres.lj)<-c("name", "intercept", "interac
t.1", "interact.2", "beta_spmk", "beta_trst", "si
gma", "epsilon")
dbres.lj[, 2:8] <- sapply(dbres.lj[, 2:8], as.num
eric)

#-----------------------------------------------
# problematics
#-----------------------------------------------
prob<-c(10, 13, 16, 17, 25, 45)
j=6

celem[,prob[j]]
#[1] "oth.ins" "price"
#-----------------------------------------------
i=prob[j]
        tmpcl1<- celem[1,i]
        tmpcl2<- celem[2,i]

        tmp<-get(paste("dblu", tmpcl1, sep=""))
        tmpcl1<-tmp[-which(tmp$type=="e"),]
        tmp<-get(paste("dblu", tmpcl2, sep=""))
        tmpcl2<-tmp[-which(tmp$type=="e"),]

#-----------------------------------------------
        tmp<-rbind(tmpcl1, tmpcl2)
        tmp[,3]<-as.character(tmp[,3])
        assign(paste("db", celem[1,i],".",celem[2,
i], sep=""), tmp)

        names(tmp)[c(1,2)]<-c("long", "lat")
        tmp[,1]<-as.numeric(as.character(tmp[,1]))
        tmp[,2]<-as.numeric(as.character(tmp[,2]))
```

```
        tmp<-unique(tmp); dim(tmp)
        tmp3<-as.data.frame(tmp[,3])
        coords<-SpatialPoints(tmp[,c("long","lat
")],proj4string=llCRS)

        tmp.e<- SpatialPointsDataFrame(coords, tmp
3) #export
        tmp.e<- spTransform(tmp.e, CRS("+proj=utm
+zone=54 ellps=WGS84")) #trsterting into mt
        assign(paste("db", celem[1,i],".",celem[2,
i],".e", sep=""), tmp.e)

        tmp.p<-as(tmp.e, "ppp")
        tmp.p<-unique.ppp(tmp.p)
        assign(paste("db", celem[1,i],".",celem[2,
i],".p", sep=""), tmp.p)

        rhat <- min(nndist(tmp.p)) #nearest neighb
or
        rhat <- rhat * 0.99999
        rhat <- ifelse(rhat==0, 0.99999, rhat)

        tmp.lj<-ppm(tmp.p, ~1, LennardJones(), cor
rection="translate", gcontrol=list(maxit=2e5), rb
ond=rhat)
        #str(tmp.lj)
        summary(tmp.lj)
        assign(paste("db", celem[1,i],".",celem[2,
i],".lj", sep=""), tmp.lj)

#tmp.lj<-aomebs.lj
tmps<-summary(tmp.lj)

dbres.lj[which(dbres.lj[,1]==as.character(dbres.l
j[prob[j], 1])),c(2:8)]<-
        c(as.numeric(coefficients(tmp.lj)),
        as.numeric(tmps$trend$value), #  beta_spmk
beta_trst
        as.numeric(tmps$interaction$sensible$print
able))

getwd()
save.image("info.Rdata")
#-----------------------------------------------

tmp.lj<-aomebs.lj
tmps<-summary(tmp.lj)

dbres.lj[1,c(2:8)]<-
        c(as.numeric(coefficients(tmp.lj)),
        as.numeric(tmps$trend$value), #  beta_spmk
beta_trst
        as.numeric(tmps$interaction$sensible$print
able))
#-----------------------------------------------
t(t(ls()[grep(".p$", ls())]))
ls()

tmp<-c("aomlubs.p","aomlubconv.p","aomlubmed.ins.
p","aomluboth.ins.p","aomlubparks.p","aomlubpric
e.p","aomlubpub.fl.p","aomlubspmk.p","aomlubtrst.
p",
"dbbus.st.conv.p","dbbus.st.med.ins.p","dbbus.st.
oth.ins.p","dbbus.st.parks.p","dbbus.st.price.p",
"dbbus.st.pub.fl.p","dbbus.st.spmk.p","dbbus.st.t
rst.p",
"dbconv.med.ins.p","dbconv.oth.ins.p","dbconv.par
ks.p","dbconv.price.p","dbconv.pub.fl.p","dbconv.
spmk.p","dbconv.trst.p",
"dbmed.ins.oth.ins.p","dbmed.ins.parks.p","dbmed.
ins.price.p","dbmed.ins.pub.fl.p","dbmed.ins.spm
k.p","dbmed.ins.trst.p",
"dboth.ins.parks.p","dboth.ins.price.p","dboth.in
s.pub.fl.p","dboth.ins.spmk.p","dboth.ins.trst.p
",
"dbparks.price.p","dbparks.pub.fl.p","dbparks.spm
k.p","dbparks.trst.p",
"dbprice.pub.fl.p","dbprice.spmk.p","dbprice.trs
t.p",
"dbpub.fl.spmk.p","dbpub.fl.trst.p",
"dbspmk.trst.p")

tmp1<-NULL
for(i in tmp)
{
        tmpf<-get(i)
        tmp1<-c(tmp1, min(nndist(tmpf)) )
}
tmp1

dbres.st$cldist<-tmp1

getwd()
write.csv(dbres.lj, "dbreslj.csv", row.names=FALS
E)
write.csv(dbres.st, "dbresst.csv", row.names=FALS
E)
#-----------------------------------------------


tmpn<-dbres.lj[,1]
tmpn1<-tmpn
tmpn1[1:9]<-gsub("aom", "", tmpn[1:9])
tmpn1[1]<-"lubus.st.lj"
#tmpn1<-gsub("lu","",tmpn1)
tmpn1<-gsub(".lj$","",tmpn1)
tmpn1<-gsub("db","",tmpn1)

cbind(tmpn, tmpn1)
celem.new<-celem
class(celem.new)

celem.names<-matrix(c("lu", "land use",
        "bus.st", "bus stops",
        "conv", "convenience stores",
        "med.ins", "medical institutions",
        "oth.ins", "other institutions",
        "parks", "parks",
        "price", "price",
        "pub.fl", "public facilities",
        "spmk", "supermarkets",
        "trst", "train stations"), ncol=2, byrow=
T)
#-----------------------------------------------
   plotLJ <- function(sigma, tau, a, b) {
        dmax <- 2 * sigma * max(1, tau)^(1/6)
        d <- seq(dmax * 0.0001, dmax, length=100
0)
        plot(d, exp(- (sigma/d)^12 + tau * (sigma
/d)^6), type="l",
                ylab="Lennard-Jones", xlab="dista
nce (m)",
                main=substitute(paste("Interrelat
ion between ", a, " and ", b, " ", sigma, " = ",
s, " ", tau, " = ", t), list(a=a, b=b, s=sigma, t
=tau) ) )
        abline(h=1, lty=2)
   }

require(lattice)

library(devtools)
for(i in 1:dim(dbres.lj)[1])
{
#       i=25
        tmp<-celem.names[match(celem[,i], celem.na
mes[,1]), 2]
        i<-ifelse(is.na(dbres.lj$sigma[i])==TRUE,
i+1, i)

        #windows(778, 639)
#       dev.size(units="in")
        pdf(file = paste("tmp",i,".pdf", sep=""),
width=12.9375, height=10.6250)
        plotLJ(dbres.lj$sigma[i], dbres.lj$epsilon
[i], a= tmp[1], b=tmp[2])
        dev.off()
}
#-----------------------------------------------
install.packages("ppls")
```

```
library(igraph)
library(ppls)

compdb<-dbres.lj
compdb[,1]<-gsub(".lj", "",compdb[,1])
compdb<-cbind(compdb, dbres.st[,c(2:5)] )

#Calculating potential
compdb$V<-4*( (compdb$sigma/compdb$cldist)^12 -
(compdb$sigma/compdb$cldist)^6 )
compdb$state<-1
compdb$state[which(compdb$V<0)]<- -1

getwd()
write.csv(compdb, "compdb.csv", row.names=F)

#extracting non complete cases
tmp<-combn(celem.names[,2], 2)
compdb.new<-cbind(type1=tmp[1,], type2=tmp[2,], c
ompdb)
compdb.new$V[which(is.na(compdb.new$V)==TRUE)]<-0

#compdb.new<-compdb[complete.cases(compdb),]
compdb.new$Vabs<-normalize.vector(abs(compdb.new
$V))*100/sd(compdb.new$V)
compdb.new$Vabs<-normalize.vector(compdb.new$Vab
s)*10

write.csv(compdb.new, "compdb_new.csv", row.names
=F)

compdb.admt<-mat.or.vec(10,10)
#upperTriangle(dbres.lj.admt, diag=TRUE) <-1
rownames(compdb.admt)<-celem.names[,1]
colnames(compdb.admt)<-celem.names[,1]
compdb.admt

head(compdb.new)
m=n=2
cont=1
for(i in 1:dim(compdb.admt)[1])
{
        for(j in n:dim(compdb.admt)[2])
        {
                compdb.admt[i,j]<-compdb.new$Vabs
[cont]
                cont=cont+1
        }
        m=n=m+1
}

compdb.admt<-round(compdb.admt, 2)
write.csv(compdb.admt, "compdb_admt.csv", row.nam
es=T)

g1 = graph.adjacency(compdb.admt, mode="undirecte
d", diag=FALSE, weighted=TRUE)
V(g1)$color<-c("red", rep("blue", length(V(g1))-
1))

plot(g1, edge.width=E(g1)$weight, vertex.size=10)
rglplot(g1, edge.width=E(g1)$weight, vertex.size=
7)
bg3d('white')
E(g1)$weight
get.edge.attribute(g1, "color")
#------------------------------------------------

#------------------------------------------------
#Starting Sendai analysis
head(senddb); dim(senddb) #[1] 30858    15

# Loading shapefiles

# Creating projection
llCRS

sendupa  <-readShapePoly("C:/Users/kuri73/Documen
ts/carlos/PhD/image processing/sendai/06/sendai06
prom.shp", proj4string=llCRS)
```

```
sendroi1<-readShapePoly("C:/Users/kuri73/Document
s/carlos/PhD/sendai city/sendma_roi1.shp", proj4s
tring=llCRS)
sendroi2<-readShapePoly("C:/Users/kuri73/Document
s/carlos/PhD/sendai city/sendma_roi2.shp", proj4s
tring=llCRS)

dim(sendupa) #[1] 31718     1

plot(sendupa, bor="gray")
unique(sendupa@data$土地利用種)
plot(sendroi2, col="red", add=T)
plot(sendroi1, col="green", add=T)

sendroi1.tar<-sapply(slot(sendroi1, "polygons"),
function(x) sapply(slot(x, "Polygons"), slot, "ar
ea"))
sendroi2.tar<-sapply(slot(sendroi2, "polygons"),
function(x) sapply(slot(x, "Polygons"), slot, "ar
ea"))
#sendroi1.tar1<-121

#-----------------------------------------------
# Starting Sendai
#-----------------------------------------------
head(senddb);dim(senddb) #[1] 30858    15
all(as.character(sendupa@data[,1]) %in% as.charac
ter(senddb$土地利用種) ) #[1] TRUE

sendupa@data$ID<-1:dim(sendupa@data)[1]
head(sendupa@data)
coordinates(sendupa)

coords<-coordinates(sendupa)
pts <- SpatialPoints(coords, proj4string=CRS(proj
4string(sendupa)))
plot(sendupa, bor="gray")
plot(pts, add=T)

pts_polys <- over(pts, sendroi1)

plot(sendupa, bor="gray")
plot(sendroi1, col=rgb(0,1,0, alpha=0.5), add=T)
plot(pts[which(is.na(pts_polys)==FALSE)], add=T,
col="red")

sendin<-pts[which(is.na(pts_polys)==FALSE)]

coords<-SpatialPoints(coordinates(sendin),proj4st
ring=llCRS)
sendin<-sendupa@data[which(is.na(pts_polys)==FALS
E),]
names(sendin)[1]<-"LU"

maindb<- SpatialPointsDataFrame(coords, sendin)
getwd()
writePointsShape(maindb,"sendccm1.shp")

maindb@data<-senddb[maindb@data$ID,]
senddb.i<-maindb@data
dim(senddb.i) #[1] 8102   15
dim(senddb.i)[1]/dim(senddb)[1] #[1] 0.2625575

#Interval
intval<-c(0, 1, 3, 5, 10, 20)

senddb.i<-senddb.i[,which(names(senddb.i)!="dothi
ns")] #inside
head(senddb.i)

tmpn<-c("dbust", "dconv", "dmedins", "dparks", "d
prc", "dpubfl", "dspmk", "dtrst")
senddb.icut<-senddb.i

for(i in tmpn)
senddb.icut[,i]<- cut(senddb.i[,i], breaks=intva
l, include.lowest=TRUE)
head(senddb.icut)

for(i in tmpn)
```

```
levels(senddb.icut[,i])<-c(1,2,3,4,5)
head(senddb.icut); dim(senddb.icut) #[1] 8102   1
5
levels(senddb.icut$土地利用種)
#-----------------------------------
# Extracting residential area

# type 7 or 7
senddb.lubs<-subset(senddb.icut, senddb.icut$土地
利用種==7)
head(senddb.lubs);dim(senddb.lubs) #[1] 5395   15

AUC.e

head(senddb.icut); dim(senddb.icut) #[1] 8102   1
5
xtable(prop.table(table(senddb.icut$土地利用種))*10
0, digits=3)
#          1           2           5           6
7          9
#0.047025426 0.020612195 0.073438657 0.003085658
0.665884967 0.042088373
#          A           B           E           F
G
#0.110466551 0.032337694 0.000000000 0.000000000
0.005060479

head(senddb.lubs)
tmp<-senddb.lubs[,c(5:12)]
dim(tmp) #[1] 3561    9
dim(tmp)[1]/dim(senddb)[1] #[1] 0.1748331

tmp1<-unique(data.matrix(unique(tmp)))
dim(tmp1) #[1] 52  8

dim(unique(senddb.icut[,c(5:12)])) #[1] 80  8
dim(tmp1)[1]/dim(unique(senddb.icut[,c(5:12)]))
[1] #[1] 0.65

ls()

#ordering according AUC
tmph<-match(AUC.e$type, colnames(tmp1))
tmph<-tmph[!is.na(tmph)]
tmp1<-tmp1[,tmph]
head(tmp1);dim(tmp1)#[1] 52  8
colnames(tmp1)

tabpt<-cbind(aucnm=as.character(AUC.e$type), lab
s)
heatmap.2(tmp1, trace="none", col=greenred(10)) #
takes time

#table(tmp1)/sum(table(tmp1))
#         1          2          3          4
#0.65384615 0.22355769 0.07211538 0.05048077

labs[tmph]
#------------------------------------------------
# Core 2
#------------------------------------------------
pts_polys2 <- over(pts, sendroi2)

plot(sendupa, bor="gray")
plot(sendroi2, col=rgb(0,1,0, alpha=0.5), add=T)
plot(pts[which(is.na(pts_polys2)==FALSE)], add=T,
col="red")

sendin2<-pts[which(is.na(pts_polys2)==FALSE)]

coords2<-SpatialPoints(coordinates(sendin2),proj4
string=llCRS)
sendin2<-sendupa@data[which(is.na(pts_polys2)==FA
LSE),]
names(sendin2)[1]<-"LU"

maindb2<- SpatialPointsDataFrame(coords2, sendin
2)
getwd()
writePointsShape(maindb2,"sendccm2.shp")
```

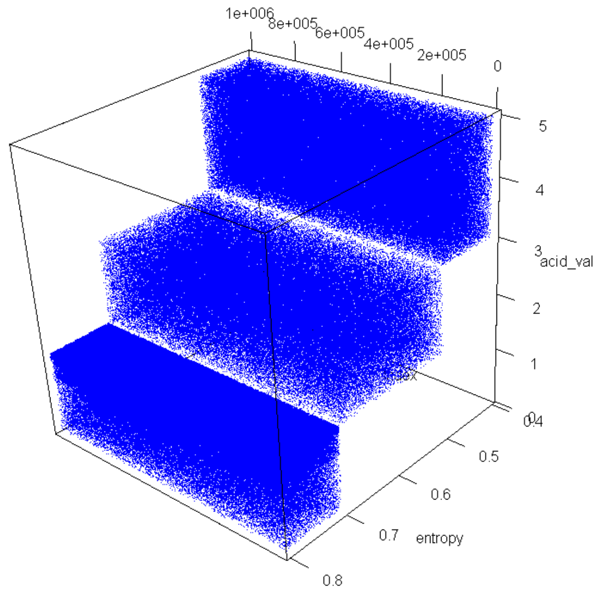
```
maindb2@data<-senddb[maindb2@data$ID,]
senddb.i2<-maindb2@data
dim(senddb.i2) #[1] 13795    15
dim(senddb.i2)[1]/dim(senddb)[1] #[1] 0.4470478

#Interval
intval<-c(0, 1, 3, 5, 10, 20)

senddb.i2<-senddb.i2[,which(names(senddb.i2)!="do
thins")] #inside
head(senddb.i2)

tmpn<-c("dbust", "dconv", "dmedins", "dparks", "d
prc", "dpubfl", "dspmk", "dtrst")
senddb.i2cut<-senddb.i2

for(i in tmpn)
senddb.i2cut[,i]<- cut(senddb.i2[,i], breaks=intv
al, include.lowest=TRUE)
head(senddb.i2cut)

for(i in tmpn)
levels(senddb.i2cut[,i])<-c(1,2,3,4,5)
head(senddb.i2cut); dim(senddb.i2cut) #[1] 8102
15
levels(senddb.i2cut$土地利用種)
#-----------------------------------
# Extracting residential area

# type 7 or 7
senddb.lubs2<-subset(senddb.i2cut, senddb.i2cut
$土地利用種==7)
head(senddb.lubs2);dim(senddb.lubs2) #[1] 9226
15

AUC.e

head(senddb.i2cut); dim(senddb.i2cut) #[1] 13795
15
xtable(prop.table(table(senddb.i2cut$土地利用種))*1
00, digits=3)
#          1           2           5           6
7          9
#0.040956868 0.017470098 0.083870968 0.002827111
0.668793041 0.039217108
#          A           B           E           F
G
#0.113591881 0.028851033 0.000000000 0.000000000
0.004421892

head(senddb.lubs2)
tmp<-senddb.lubs2[,c(5:12)]
dim(tmp) #[1] 9226    8
dim(tmp)[1]/dim(senddb)[1] #[1] 0.2989824

tmp1<-unique(data.matrix(unique(tmp)))
dim(tmp1) #[1] 66  8

dim(unique(senddb.i2cut[,c(5:12)])) #[1] 102   8
dim(tmp1)[1]/dim(unique(senddb.i2cut[,c(5:12)]))
[1] #[1] 0.6470588

#ordering according AUC
tmph<-match(AUC.e$type, colnames(tmp1))
tmph<-tmph[!is.na(tmph)]
tmp1<-tmp1[,tmph]
head(tmp1);dim(tmp1)#[1] 66  8
colnames(tmp1)

tabpt<-cbind(aucnm=as.character(AUC.e$type), lab
s)
heatmap.2(tmp1, trace="none", col=greenred(10)) #
takes time

#table(tmp1)/sum(table(tmp1))
#          1           2           3           4
5
#0.662878788 0.214015152 0.062500000 0.054924242
0.005681818
```

```
#-------------------------------------------------

#-------------------------------------------------
#Starting Sapporo analysis
head(sapdb); dim(sapdb) #[1] 72745    15
summary(sapdb)

# Loading shapefiles

# Creating projection
llCRS

sapupa  <-readShapePoly("C:/Users/kuri73/Documents/carlos/PhD/image processing/sapporo/06/sapporo06prom.shp", proj4string=llCRS)
saproi1<-readShapePoly("C:/Users/kuri73/Documents/carlos/PhD/sapporo city/sapma_roi1.shp", proj4string=llCRS)
saproi2<-readShapePoly("C:/Users/kuri73/Documents/carlos/PhD/sapporo city/sapma_roi2.shp", proj4string=llCRS)

dim(sapupa) #[1] 31718     1

plot(sapupa, bor="gray")
unique(sapupa@data$土地利用種)
plot(saproi2, col="red", add=T)
plot(saproi1, col="green", add=T)

saproi1.tar<-sapply(slot(saproi1, "polygons"), function(x) sapply(slot(x, "Polygons"), slot, "area"))
saproi2.tar<-sapply(slot(saproi2, "polygons"), function(x) sapply(slot(x, "Polygons"), slot, "area"))
#saproi1.tar1<-121
#-------------------------------------------------

#-------------------------------------------------
# Starting Sapporo
#-------------------------------------------------
head(sapdb); dim(sapdb) #[1] 72745    15
summary(sapdb)

all(as.character(sapupa@data[,1]) %in% as.character(sapdb$土地利用種) ) #[1] TRUE

sapupa@data$ID<-1:dim(sapupa@data)[1]
head(sapupa@data)
#coordinates(sapupa)

coords<-coordinates(sapupa)
pts <- SpatialPoints(coords, proj4string=CRS(proj4string(sapupa)))
plot(sapupa, bor="gray")
plot(pts, add=T)

pts_polys <- over(pts, saproi1)

plot(sapupa, bor="gray")
plot(saproi1, col=rgb(0,1,0, alpha=0.5), add=T)
plot(pts[which(is.na(pts_polys)==FALSE)], add=T, col="red")

sapin<-pts[which(is.na(pts_polys)==FALSE)]

coords<-SpatialPoints(coordinates(sapin),proj4string=llCRS)
sapin<-sapupa@data[which(is.na(pts_polys)==FALSE),]
names(sapin)[1]<-"LU"

maindb<- SpatialPointsDataFrame(coords, sapin)
getwd()
writePointsShape(maindb,"sapccm1.shp")

maindb@data<-sapdb[maindb@data$ID,]
sapdb.i<-maindb@data
dim(sapdb.i) #[1] 20991    15
dim(sapdb.i)[1]/dim(sapdb)[1] #[1] 0.2885559

#Interval
intval<-c(0, 1, 3, 5, 10, 20)

sapdb.i<-sapdb.i[,which(names(sapdb.i)!="dothins")] #inside
head(sapdb.i)

tmpn<-c("dbust", "dconv", "dmedins", "dparks", "dprc", "dpubfl", "dspmk", "dtrst")
sapdb.icut<-sapdb.i

for(i in tmpn)
sapdb.icut[,i]<- cut(sapdb.i[,i], breaks=intval, include.lowest=TRUE)
head(sapdb.icut)

for(i in tmpn)
levels(sapdb.icut[,i])<-c(1,2,3,4,5)
head(sapdb.icut); dim(sapdb.icut) #[1] 20991    15
levels(sapdb.icut$土地利用種)
#-----------------------------------
# Extracting residential area

# type 7 or 7
sapdb.lubs<-subset(sapdb.icut, sapdb.icut$土地利用種==7)
head(sapdb.lubs);dim(sapdb.lubs) #[1] 14418    15

head(sapdb.icut); dim(sapdb.icut) #[1] 20991    15
options(digits=6)
round(prop.table(table(sapdb.icut$土地利用種)), 6)
xtable(prop.table(table(sapdb.icut$土地利用種))*100, digits=3)
#       1        2        5        6        7        9        A        B
#0.000810 0.035777 0.041637 0.012434 0.686866 0.047115 0.152494 0.021247
#       E        F        G
#0.000095 0.000000 0.001524

head(sapdb.lubs)
tmp<-sapdb.lubs[,c(5:12)]
dim(tmp) #[1] 14418     8
dim(tmp)[1]/dim(sapdb)[1] #[1] 0.198199

tmp1<-unique(data.matrix(unique(tmp)))
dim(tmp1) #[1] 63  8

dim(unique(sapdb.icut[,c(5:12)])) #[1] 99  8
dim(tmp1)[1]/dim(unique(sapdb.icut[,c(5:12)]))[1] #[1] 0.636364

#ordering according AUC
tmph<-match(AUC.e$type, colnames(tmp1))
tmph<-tmph[!is.na(tmph)]
tmp1<-tmp1[,tmph]
head(tmp1);dim(tmp1)#[1] 63  8
colnames(tmp1)

tabpt<-cbind(aucnm=as.character(AUC.e$type), labs)
heatmap.2(tmp1, trace="none", col=greenred(10)) # takes time

#table(tmp1)/sum(table(tmp1))
#        1         2         3         4
#0.6190476 0.2638889 0.0773810 0.0396825
#-------------------------------------------------

#-------------------------------------------------
# Core 2
#-------------------------------------------------
pts_polys2 <- over(pts, saproi2)

plot(sapupa, bor="gray")
plot(saproi2, col=rgb(0,1,0, alpha=0.5), add=T)
plot(pts[which(is.na(pts_polys2)==FALSE)], add=T,
```

```
col="red")

sapin2<-pts[which(is.na(pts_polys2)==FALSE)]

coords2<-SpatialPoints(coordinates(sapin2),proj4s
tring=llCRS)
sapin2<-sapupa@data[which(is.na(pts_polys2)==FALS
E),]
names(sapin2)[1]<-"LU"

maindb2<- SpatialPointsDataFrame(coords2, sapin2)
getwd()
writePointsShape(maindb2,"sapccm2.shp")

maindb2@data<-sapdb[maindb2@data$ID,]
sapdb.i2<-maindb2@data
dim(sapdb.i2) #[1] 31466    15
dim(sapdb.i2)[1]/dim(sapdb)[1] #[1] 0.432552

#Interval
intval<-c(0, 1, 3, 5, 10, 20)

sapdb.i2<-sapdb.i2[,which(names(sapdb.i2)!="dothi
ns")] #inside
head(sapdb.i2)

tmpn<-c("dbust", "dconv", "dmedins", "dparks", "d
prc", "dpubfl", "dspmk", "dtrst")
sapdb.i2cut<-sapdb.i2

for(i in tmpn)
sapdb.i2cut[,i]<- cut(sapdb.i2[,i], breaks=intva
l, include.lowest=TRUE)
head(sapdb.i2cut)

for(i in tmpn)
levels(sapdb.i2cut[,i])<-c(1,2,3,4,5)
head(sapdb.i2cut); dim(sapdb.i2cut) #[1] 8102   1
5
levels(sapdb.i2cut$土地利用種)
#-----------------------------------
# Extracting residential area

# type 7 or 7
sapdb.lubs2<-subset(sapdb.i2cut, sapdb.i2cut$土地
利用種==7)
head(sapdb.lubs2);dim(sapdb.lubs2) #[1] 20397
15

head(sapdb.i2cut); dim(sapdb.i2cut) #[1] 31466
15
xtable(prop.table(table(sapdb.i2cut$土地利用種))*10
0, digits=3)
#          1           2           5           6
7          9
#0.002033941 0.043380156 0.046939554 0.017542745
0.648223479 0.042522087
#          A           B           E           F
G
#0.173012140 0.021642408 0.000572046 0.002065722
0.002065722

head(sapdb.lubs2)
tmp<-sapdb.lubs2[,c(5:12)]
dim(tmp) #[1] 20397     8
dim(tmp)[1]/dim(sapdb)[1] #[1] 0.28039

tmp1<-unique(data.matrix(unique(tmp)))
dim(tmp1) #[1] 98  8

dim(unique(sapdb.i2cut[,c(5:12)])) #[1] 163   8
dim(tmp1)[1]/dim(unique(sapdb.i2cut[,c(5:12)]))
[1] #[1] 0.601227

#ordering according AUC
tmph<-match(AUC.e$type, colnames(tmp1))
tmph<-tmph[!is.na(tmph)]
tmp1<-tmp1[,tmph]
head(tmp1);dim(tmp1)#[1] 98  8
colnames(tmp1)

tabpt<-cbind(aucnm=as.character(AUC.e$type), lab
s)
heatmap.2(tmp1, trace="none", col=greenred(10)) #
takes time

#table(tmp1)/sum(table(tmp1))
#        1         2         3         4
#0.5420918 0.2959184 0.0816327 0.0803571
#-------------------------------------------------

#-------------------------------------------------
# Saving
#-------------------------------------------------
setwd(maindir)
getwd()
options(encoding="SHIFT-JIS")
save.image("info.Rdata")
```

SIMULATING UPAs

```
library(rgl)

nsim<-1000000
core<-1

area <-as.integer(runif(nsim, 4000, 30000))

entr <-ifelse(area<6000, runif(nsim, 0.05, 1),
       ifelse(area<10000, runif(nsim, 1, 1.4),
       ifelse(area<20000, runif(nsim, 1.4, 1.5),
runif(nsim, 1.5, 5) ) )) #entropy

nvect<- as.integer(ifelse(area<8000, runif(nsim,
40, 47),
                      ifelse(area<10000,
runif(nsim, 47, 50),
                              ifelse(area<20000,
runif(nsim, 50, 60), runif(nsim, 60, 100)) )) )
#number of vectors

perar<-ifelse(entr<1.5, runif(nsim, 0.7, 0.8),
       ifelse(entr<3, runif(nsim, 0.5, 0.7),
runif(nsim, 0.4, 0.5) ) ) #percentage


det.area<-ifelse(perar<0.3,
as.integer(runif(nsim, 3, 4)),
        ifelse(perar<0.8, as.integer(runif(nsim,
1, 2)), 0) )

#new compact value calculated
new.cmpct<-1/(core + det.area)

db<-cbind(area, nvect, entr, perar, det.area,
new.cmpct)
db<-as.data.frame(db)
db[sample(nrow(db), 10),]

#----------------------------
# p percentage
# a area
# e entropy
# nc new compact
# nv number of vectors

fcalc2<-function(p, a, e, nc, nv)
{
tmp<-p * log(a * exp(e * nc/nv) )
return(tmp)
}

aom1<-fcalc2(.87, 5261, 1.286, 1.00, 40) #[1]
7.482197
snd1<-fcalc2(.28, 8746, 1.411, 0.33, 52) #[1]
2.543886
sap1<-fcalc2(.32, 23625, 1.559, 0.20, 72) #[1]
3.223805

aom2<-fcalc2(1, 5261, 1.286, 1.00, 40) #[1]
7.482197
snd2<-fcalc2(1, 8746, 1.411, 0.33, 52) #[1]
2.543886
sap2<-fcalc2(1, 23625, 1.559, 0.20, 72) #[1]
3.223805

aom.bv<-fcalc2(1, 5261, .08, 1, 40) #[1] 8.570076
snd.bv<-fcalc2(1, 8746, .08, 1, 40) #[1] 9.078352
sap.bv<-fcalc2(1, 23625, .08, 1, 40) #[1] 10.07206

aom1/aom.bv
snd1/snd.bv
sap1/sap.bv

aom2/aom.bv
snd2/snd.bv
sap2/sap.bv

head(db)
db$calc<-fcalc2(db$perar, db$area, db$entr,
db$new.cmpct, db$nvect)

db[sample(1:nrow(db), 50,replace=FALSE),]

plot3d(db$perar, db$entr, db$calc1, col=rgb(0,0,1,
0.1), size=.5 )
#plot3d(db$perar, db$nvect, db$entr,
col=rgb(0,0,1, 0.1), size=.5 )

#------------------------------
#maximum value
tmpmax<-db[which(db$calc==max(db$calc)),]

#minimum value
tmpmin<-db[which(db$calc==min(db$calc)),]
tmpmin
#------------------------------
tmpmax$calc/fcalc2(1, tmpmax$area, 1, 1, 40)

head(db)
tail(sort(table(db$area)))

tmp<-db[which(db$area==19911),]

#------------------------------

tmpp<- seq(0,1, length.out=100)
tmpe<- seq(1,4, length.out=100)
tmpnv<-1:tmpmin$nvect
tmp<-expand.grid(tmpp, tmpe)

newdb<-NULL
cont<-1
for(j in tmpp)
{
for(i in tmpnv)
{
        newdb<-rbind(newdb, fcalc2(j, tmpmin$area,
tmpmin$entr, tmpmin$new.cmpct, i))
}
assign(paste("newdb", "_", cont, sep=""), newdb)
newdb<-NULL
cont<-cont+1
}

ls()
#rm(list=ls()[grep("newdb", ls())])

plot(get(paste("newdb", "_",2,sep="")), type="l",
xlim=c(0,tmpmin$nvect), ylim=c(0,15))

for(i in 3:100)
{
#i=3
lines(get(paste("newdb", "_",i,sep="")))
}

#-----------------------------------------------
-----
sort(table(area))

tmp<-which(table(area)==max(table(area)))
tmpdb<-
db[which(db$area==as.numeric(names(tmp[1]))), ]
tmpdb

setwd("C:/Users/kuri73/Documents/carlos/PhD/expe
riments/last part")
save.image("sim_newt.Rdata")
```


1e+006
8e+005
6e+005
4e+005
2e+005
0
5
4
3
acid_val
2
1
0
0.4
0.5
0.6
0.7
entropy
0.8

BIBLIOGRAPHY

[1] *OECD Green Growth Studies Compact City Policies A Comparative Assessment: A Comparative Assessment*. OECD Green Growth Studies. OECD Publishing, 2012. ISBN 9789264167865. URL http://books.google.co.jp/books?id=fI6lrYZABkcC.

[2] P. Agouris and A. Croituru. *Next Generation Geospatial Information: From Digital Image Analysis to Spatiotemporal Databases*. ISPRS Book Series. Taylor & Francis, 2005. ISBN 9780415380492. URL http://books.google.co.jp/books?id=5qoh5FwFCs4C.

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Proceedings of the 4th International Conference on Ambient Assisted Living and Home Care*, IWAAL'12, pages 216–223, Berlin, Heidelberg, 2012. Springer-Verlag.

[4] Shigeru Araki. Change in population and land-use intensities in several villages of the four northern regions of Namibia. *African study monographs*, 30(March):77–88, 2005.

[5] Adrian Baddeley. Spatial point patterns: Models and statistics. In *Stochastic Geometry, Spatial Statistics and Random Fields*, pages 49–114. Springer Berlin Heidelberg, 2013.

[6] Adrian Baddeley and Rolf Turner. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005. URL http://www.jstatsoft.org/v12/i06/.

[7] Adrian Baddeley, Rolf Turner, Jorge Mateu, and Andrew Bevan. Hybrids of gibbs point process models and their implementation. *Journal of Statistical Software*, 55(11):1–43, 2013. URL http://www.jstatsoft.org/v55/i11/.

[8] Roger S. Bivand, Edzer Pebesma, and Virgilio Gomez-Rubio. *Applied spatial data analysis with R, Second edition*. Springer, NY, 2013. URL http://www.asdar-book.org/.

[9] Ulf Brefeld and Tobias Scheffer. Auc maximizing support vector learning. In *Proceedings of the ICML 2005 workshop on ROC Analysis in Machine Learning*. Citeseer, 2005.

[10] Elizabeth Burton. The Compact City and Social Justice. In *Housing, Environment and sustainability*, number April, pages 1–16, York, 2001.

[11] Elizabeth Burton. Measuring urban compactness in UK towns and cities. *Environment and Planning B: Planning and Design*, 29(2):219–250, 2002. URL http://ideas.repec.org/a/pio/envirb/v29y2002i2p219-250.html.

[12] Peter Calthorpe and William Fulton. *The regional city*. Island Press, 2001.

[13] Colin Clark. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)*, 114(4):pp. 490–496, 1951. ISSN 00359238. URL http://www.jstor.org/stable/2981088.

[14] L. Corbusier, P. Jeanneret, and Fondation Le Corbusier. *Urbanisme, Algiers, and other buildings and projects, 1930-1933*. Le Corbusier archive. Garland, 1983.

[15] G.B. Dantzig and T.L. Saaty. *Compact City: A Plan for a Liveable Urban Environment*. W. H. Freeman, 1973.

[16] C. Denpaiboon and Kanegae H. Compact city strategy of bangkok mega-city. *Journal of Humanities*, 11(2):38–55, November 2008.

[17] Howard DeVoe. *Thermodynamics and Chemistry*. Prentice Hall, 2nd edition, 2012. URL http://www2.chem.umd.edu/thermobook/.

[18] Izabela Dyras, Hartwig Dobesch, Estelle Grueter, Antonio Perdigao, Ole E Tveito, John E Thornes, Frans van der Wel, and Lorenzo Bottai. The use of geographic information systems in climatology and meteorology: Cost 719. *Meteorological Applications*, 12(1):1–5, 2005.

[19] Mageret Ely. Urban rail networks in world cities. *Journeys*, 8: 48–57, 2012.

[20] Reid Ewing. Is los angeles-style sprawl desirable? *Journal of the American planning association*, 63(1):107–126, 1997.

[21] David Felce and Jonathan Perry. Quality of life: Its definition and measurement. *Research in Developmental Disabilities*, 16(1): 51 – 74, 1995. ISSN 0891-4222. doi: http://dx.doi.org/10.1016/0891-4222(94)00028-8. URL http://www.sciencedirect.com/science/article/pii/0891422294000288.

[22] Giles M Foody and Ajay Mathur. Toward intelligent training of supervised image classifications: directing training data acquisition for svm classification. *Remote Sensing of Environment*, 93 (1):107–117, 2004.

[23] L.J. Frewer, E. Risvik, and H. Schifferstein. *Food, People and Society: A European Perspective of Consumers' Food Choices*. Springer, 2001. ISBN 9783540415213. URL http://books.google.co.jp/books?id=RyYM4OuSixUC.

[24] Yoshiharu Fukuda, Keiko Nakamura, and Takehito Takano. Municipal socioeconomic status and mortality in Japan: sex and age differences, and trends in 1973-1998. *Social science & medicine*, 59(12):2435–2445, December 2004.

[25] Yoshiharu Fukuda, Keiko Nakamura, and Takehito Takano. Wide range of socioeconomic factors associated with mortality among cities in Japan. *Health Promotion International*, 19(2): 177–187, June 2004.

[26] T. Fukumura. *The Omiya Sonic City–a Case Study of the Public-private Partnership in Japan*. Massachusetts Institute of Technology, Department of Urban Studies and Planning, 1994.

[27] Raghav Gaiha and Ganesh Thapa. Supermarkets, smallholders and livelihood prospects in selected asian countries. *Occasional Papers: Knowledge for development effectiveness*, 4, 2007.

[28] Will Galloway. Of suburbs and cities. *Town and country planning*, pages 322–327, 2009.

[29] George Galster, Royce Hanson, Michael R Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage. Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing policy debate*, 12(4):681–717, 2001.

[30] P. Goovaerts. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228(1):113 – 129, 2000.

[31] Peter Gordon and Harry W Richardson. Are compact cities a desirable planning goal? *Journal of the american planning association*, 63(1):95–106, 1997.

[32] C. A. Gotway and W. W. Stroup. A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(2):pp. 157–178, 1997. ISSN 10857117. URL http://www.jstor.org/stable/1400401.

[33] Hideo HARASAWA. Compact city project. *C]. Proceedings of IGES/APN Mega-City Project (Rihga Royal Hotel Kokura, Kitakyushu Japan)*, pages 23–25, 2002.

[34] Graham Haughton and Colin Hunter. *Sustainable Cities*. Regional development and public policy. Routledge, 2003.

[35] K.E. Haynes and A.S. Fotheringham. *Gravity and Spatial Interaction Models*. Scientific Geography Series. SAGE Publications, 1984. ISBN 9780803923263. URL http://books.google.co.jp/books?id=guTrAAAAMAAJ.

[36] T. Hengl. *A Practical Guide to Geostatistical Mapping of Environmental Variables*. EUR 22904 EN - 2007. European Commission – Joint Research Centre Institute for Environment and Sustainability, 2nd edition, 2007.

[37] Petr Honzik, Pavel Kucera, Ondrej Hyncica, and Vaclav Jirsik. Novel Method for Evaluation of Multi-class Area Under Receiver Operating Characteristic. In *Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009. ICSCCW 2009. Fifth International Conference on*, pages 1–4, Famagusta, 2009.

[38] C. Huang, L.S. Davis, and J.R.G Townshend. An assessment of support vector machines for land cover classication. *International journal of remote sensing*, 23(4):725–749, 2002.

[39] David L Huff. A probabilistic analysis of shopping center trade areas. *Land economics*, pages 81–90, 1963.

[40] Maarten Jansen, Maurits Malfait, and Adhemar Bultheel. Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):33–44, 1997.

[41] J. E. Jones. On the determination of molecular fields. i. from the variation of the viscosity of a gas with temperature. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):pp. 441–462, 1924. ISSN 09501207.

[42] Narufumi Kadomatsu. Recent Development of Decentralization, Deregulation and Citizens' Participation in Japanese City planning Law. *Kobe University Law Review*, 40:1–14, 2006.

[43] Alexandros Karatzoglou, Alex Smola, Kurt Hornik, and Achim Zeileis. kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9):1–20, 2004. URL http://www.jstatsoft.org/v11/i09/.

[44] Yukichika Kawata. Socioeconomic Factors Influencing Longevity in Japan. *Atlantic Economic Journal*, 37(1):113–114, October 2008.

[45] D. Klimesova and E. Ocelikova. Spatial data modelling and maximum entropy theory. *Agricultural Economics*, 51(2):80–83, 2005.

[46] Jan Knorn, Andreas Rabe, Volker C. Radeloff, Tobias Kuemmerle, Jacek Kozak, and Patrick Hostert. Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment*, 118:957–964, May 2009.

[47] Takeshi Kobayashi, Shinji Ikaruga, and Mahito Nakazono. Study on forecasting urbanization by applying area division system to the suburb in local cities. In *Proc. 9th International Conference on Computers in Urban Planning and Urban Management*, pages 1–15, London (England), June 2005.

[48] N. S.-N. Lam. Spatial interpolation methods: a review. *The American Cartographer*, 10:129–149, 1983.

[49] Michael Lavine. *Introduction to Statistical Thought*. 2010. URL http://www.math.umass.edu/~lavine/Book/book.pdf.

[50] X. H. Liu, P. C. Kyriakidis, and M. F. Goodchild. Population - density estimation using regression and area to point residual kriging. *International Journal of Geographical Information Science*, 22(4):431–447, apr 2008.

[51] X.H. Liu. *Estimation of the Spatial Distribution of Urban Population Using High Spatial Resolution Satellite Imagery*. University of California, Santa Barbara, 2003. URL http://books.google.co.jp/books?id=a98POAAACAAJ.

[52] MG Logsdon and EJ Bell. Probability mapping of land use change: A GIS interface for visualizing transition probabilities. *Computers, environment and urban*, 20(6):389–398, 1996.

[53] Luis C. Manrique R. and Kayoko Yamamoto. Support vector machine applied to land use prediction using socio-economic factors in a compact city model. *International Journal Of Mathematical Models And Methods In Applied Sciences*, 7(1):945–955, 2013.

[54] Luis C. Manrique R. and Kayoko Yamamoto. Compact city model analysis for a metropolitan area using gis: Relationship between population and land use. *International journal of applied mathematics and informatics*, 8:82–92, 2014.

[55] P.J. Marcotullio. The compact city, environmental transition theory and asia-pacific urban sustainable development. In *International workshop new approaches to land management for sustainable urban regions*, pages 1–25, 2001.

[56] David Martin. An assessment of surface and zonal models of population. *International journal of geographical information systems*, 10(8):973–989, 1996. doi: 10.1080/

02693799608902120. URL http://www.tandfonline.com/doi/abs/10.1080/02693799608902120.

[57] Shigeru Mase. Mean characteristics of gibbsian point processes. *Annals of the Institute of Statistical Mathematics*, 42(2):203–220, 1990.

[58] A. Mathur and G.M. Foody. Multiclass and binary SVM classification: Implications for training and classification users. *Geoscience and Remote Sensing Letters, IEEE*, 5(2):241–245, April 2008.

[59] James J. McCarthy, Osvaldo F. Canziani, Neil A. Leary, David J. Dokken, and Kasey S. White. *Climate Change 2001: Impacts, Adaptation and Vulnerability*. Ed. Cambridge University Press., 2001. URL http://www.grida.no/publications/other/ipcc_tar/?src=/climate/ipcc_tar/wg2/index.htm.

[60] Harvey J. Miller and Jiawei Han. *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, Inc., Bristol, PA, USA, 2009.

[61] I.L. Muntean and L.M. Dansorean. Searching simulation scenarios on the grid with elsigexplorer. In *Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2011 13th International Symposium on*, pages 261–267, Sept 2011.

[62] Gordon Frank Newell. *Traffic flow on transportation networks*. Number Monograph. 1980.

[63] United States. Advisory Commission on Intergovernmental Relations. *Metropolitan social and economic disparities: implications for intergovernmental relations in central cities and suburbs*. A (Estados Unidos. Advisory Commission on Intergovernmental Relations). The Commission, 1965. URL http://books.google.co.jp/books?id=uWKShqdKUEkC.

[64] Thomas Oommen, Debasmita Misra, Navin K. C. Twarakavi, Anupma Prakash, Bhaskar Sahoo, and Sukumar Bandopadhyay. An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing. *Mathematical Geosciences*, 40(4):409–424, March 2008.

[65] Sean O'Sullivan and John Morrall. Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*, 1538(1):19–26, 1996.

[66] Edzer J. Pebesma. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30:683–691, 2004.

[67] Mats Pemer. Developing A Sustainable Compact City in Stockholm, Sweden. Thematic Committee. Technical report, United

Nations Centre for Human Settlements (Habitat), New York, 2001.

[68] Duane Perry, S Sherman, M Stone, A Hillier, and G Pozoukidou. Food for every child: the need for more supermarkets in philadelphia. *Farmers' Market Trust*, 2002.

[69] Prasan Pitiranggon, Somsri Banditvilai, and Nunthika Benjathepanun. Detection of Currency Crises by a Novel Rule Extraction Method from Support Vector Machine. *International journal of mathematical models and methods in applied sciences*, 4(3): 141–149, 2010.

[70] Antonio Plaza, Jon Atli Benediktsson, Joseph W Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, et al. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113:S110–S122, 2009.

[71] Foster Provost and Tom Fawcett. Robust classification for imprecise environments. *Machine Learning*, 42(3):203–231, 2001.

[72] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL http://www.R-project.org/. ISBN 3-900051-07-0.

[73] TV Ramachandra, Bharath H Aithal, and S Vinay. Comprehension of temporal land use dynamics in urbanising landscape. In *Proceedings of National Remote Sensing Centre, ISRO*, pages 1–6, Balanagar, Hyderabad, India, February 2013.

[74] Kelsey Ramírez-Gutiérrez, Daniel Cruz-Pérez, and Héctor Pérez-Meana. Face recognition and verification using histogram equalization. In *Proceedings of the 10th WSEAS International Conference on Applied Computer Science*, ACS'10, pages 85–89, Stevens Point, Wisconsin, USA, 2010. World Scientific and Engineering Academy and Society (WSEAS).

[75] Kelsey Ramirez-Gutierrez, Daniel Cruz-Perez, Jesús Olivares-Mercado, Mariko Nakano-Miyatake, and Hector Perez-Meana. A face recognition algorithm using eigenphases and histogram equalization. *International journal of Computers*, 5(1):34–41, 2011.

[76] Carlo Ratti and Paul Richens. Raster analysis of urban form. *Environment and Planning B: Planning and Design*, 31(2):297–309, 2004.

[77] C.T. Reviews. *e-Study Guide for Cities of the World: World Regional Urban Development, textbook by Stanley D. Brunn: Earth sciences, Physical geography*. Cram101, 2013. ISBN 9781478458036. URL http://books.google.co.jp/books?id=n8ChoXx-4dMC.

[78] Adriana Reyes, Ramón Giraldo, and Jorge Mateu. Residual kriging for functional data. application to the spatial prediction of salinity curves (available online). *Communications in Statistics. Theory and Methods*, 2012. URL http://www.icn.unal.edu.co/unciencias/data-file/user_45/RI20.pdf.

[79] Muhammad Sani Roychansyah, Koichi Ishizaka, and Takashi Omi. Considerations of regional characteristics for delivering city compactness:. *Journal of Asian Architecture and Building Engineering*, 4(2):339–346, 2005.

[80] Jan Scheurer. *Urban Ecology, Innovations in Housing Policy and the Future of Cities: Towards Sustainability in Neighbourhood Communities*. Number Vcd. Murdoch University., 2001. URL http://www.istp.murdoch.edu.au/ISTP/publications/jscheurer/urbanecology/.

[81] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[82] Sulochana Shekhar. Urban sprawl assessment entropy approach. *GIS Development, Noida*, 2004.

[83] B. Shelton. *Learning from the Japanese City: Looking East in Urban Design*. Planning, History and Environment Series. Taylor & Francis, 2012.

[84] Dee Shi and Xiaojun Yang. Support Vector Machines for Landscape Mapping from Remote Sensor Imagery. In *AutoCarto 2012*, pages 1–12, Columbus, Ohio, USA, 2012.

[85] David Marshall Smith. *Patterns in human geography*. Penguin Harmondsworth, 1977.

[86] G.D. Squires. *Urban Sprawl: Causes, Consequences, & Policy Responses*. Urban Institute Press, 2002.

[87] Balaji Vasan Srinivasan, Ramani Duraiswami, and Raghu Murtugudde. Efficient kriging for real-time spatio-temporal interpolation Linear kriging. In *20th conference on probability and statistics in the atmospheric sciences*, pages 228–235, Atlanta, 2010.

[88] M.L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York, 1999. ISBN 9780387986296. URL http://books.google.co.jp/books?id=5n_XuL2Wx1EC.

[89] John Q. Stewart and William Warntz. Physics of population distribution. *Journal of Regional Science*, 1(1):99–121, 1958. ISSN 1467-9787. doi: 10.1111/j.1467-9787.1958.tb01366.x. URL http://dx.doi.org/10.1111/j.1467-9787.1958.tb01366.x.

[90] L R Sudha and R Bhavani. Performance Comparison of SVM and kNN in Automatic Classification of Human Gait Patterns. *International Journal of computers*, 6(1):19–28, 2012.

[91] Heng Sun, Wayne Forsythe, and Nigel Waters. Modeling urban land use change and urban sprawl: Calgary, alberta, canada. *Networks and spatial economics*, 7(4):353–376, 2007.

[92] S. Szalai, Z. Bihari, T. Szentimrey, and M. Lakatos, editors. *COST Action 719 — The Use of Geographic Information Systems in Climatology and Meteorology*. Office for Official Publications of the European Communities, Luxemburg, 2007.

[93] Yu-Hsin Tsai. Quantifying urban form: compactness versus' sprawl'. *Urban studies*, 42(1):141–161, 2005.

[94] Margaret Walls. Parks and recreation in the united states: State park systems. *Washington, DC: Resources for the Future*, 2009.

[95] G. Woodman. *Cultural Shock-Taiwan*. Xlibris Corporation, 2010. ISBN 9781462813971. URL http://books.google.co.jp/books?id=AYAHTAntZuEC.

[96] Shuo-sheng Wu, Xiaomin Qiu, and Le Wang. Population estimation methods in gis and remote sensing: a review. *GIScience & Remote Sensing*, 42(1):80–96, 2005.

[97] Kayoko Yamamoto. Evaluation of Public Green Space Placement Plans as an Indicator of Urban Density of Japan's Three Major Metropolitan Areas. In *Proceedings of Joint international symposium and exhibition on geoinformation 2007 and international symposium on GPS/GNSS 2007. Bharu*, pages 42–53, Malaysia nov. 5-7, 2007.

[98] Kayoko Yamamoto. Evaluation of the Degree of the Sufficiency of Public Green Spaces as an Indicator of Urban Density in Metropolitan Areas in Japan: The Chubu Metropolitan Area. *International Journal of Advanced Computer Science*, 2(4):168–180, 2012.

[99] Kayoko Yamamoto and Hideharu Morishita. An evaluation of the locations of open spaces as regional refuge using geographical information systems (gis): A case study of tokyo. *Papers on environmental information science*, 12:125–130, nov 1998. ISSN 03896633.

[100] Kayoko Yamamoto and Masahisa Nakamura. An examination of land use controls in the Lake Biwa watershed from the perspective of environmental conservation and management. *Lakes and Reservoirs: Research and Management*, 9(3-4):217–228, September 2004.

[101] Yizhao Yang. A tale of two cities: Physical form and neighborhood satisfaction in metropolitan portland and charlotte. *Journal of the American Planning Association*, 74(3):307–323, 2008.

[102] Hwanjo Yu. Svm selective sampling for ranking with application to data retrieval. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, pages 354–363, New York, NY, USA, 2005. ACM. ISBN 1-59593-135-X. doi: 10.1145/1081870.1081911. URL http://doi.acm.org/10.1145/1081870.1081911.

[103] Huiwei Zhou, Tong Yu, Degen Huang, A Support Vector, and Machine Svm. Japanese Dependency Analysis Based on SVMs and CRFs. *International journal of mathematics and computers in simulation*, 1(3):233–237, 2007.

[104] Guobin Zhu and Dan G Blumberg. Classification using aster data and svm algorithms;: The case study of beer sheva, israel. *Remote sensing of Environment*, 80(2):233–240, 2002.

PUBLICATIONS

Journal Papers

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Support vector machine for land use through socio-economic factors applied to a compact city model. International Journal of Mathematical Models and Methods in Applied Sciences. Vol.7, No.11, 2013, pp.945-955.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Compact City Model Analysis for a Metropolitan Area using GIS, Relationship between Population and Land Use. International Journal of Applied Mathematics and Informatics. Vol. 8, 2014, pp.82-92.

Proceedings

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Support vector machine for land use through socio-economic factors applied to a compact city model. Proceedings of the 12th International Conference on System Science and Simulation In Engineering (ICOSSSE '13). Recent Advances in Automatic Control, Modeling & Simulation. No.12, 2013, pp.113-120. Peer reviewed.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Classification Methods of Socio-Economic Factors Applied in a Compact City Model. Annual Meeting Proceedings of JAPA & SSI Youth Conference 2013. 2013, pp.95-98.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Spatial Modeling Applied to the Cities in the Northern and eastern areas of Japan. Annual Meeting Proceedings of JAPA & SSI Youth Conference 2012. 2012, pp.87-90.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Spatial Modeling Applied to the Cities in Japan. Annual Meeting Proceedings of Japan Geoscience Union Meeting 2012. 2012, 2p, (CD-ROM).

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Transition probabilities applied to land use changes in the northern metropolitan areas of Japan. Geographic Information Systems Association Papers and Proceedings. Vol.20, 2012, 4p, (CD-ROM).

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Compact City Analysis in Northern and Eastern Areas of Japan. Geographic Information Systems Association Papers and Proceedings. Vol.19, 2011, 6p, (CD-ROM).

International conferences

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Support vector machine for land use through socio-economic factors applied to a compact city model. 12th International Conference on System Science and Simulation In Engineering (ICOSSSE '13). Morioka Malios, 2013.

Domestic conferences

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Classification methods of socio-economic factors applied in a compact city model. JAPA & SSI Youth conference 2013. Meisei University, 2013.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Transition probabilities applied to land use changes in the northern metropolitan areas of Japan. 21th Annual Conference of GIS Association of Japan. Hiroshima Shudo University, 2012.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Spatial Modeling Applied to the Cities in Japan. Annual Meeting of Japan Geoscience Union Meeting 2012. Makuhari Messe, 2012.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Spatial modeling applied to the cities in Japan. JAPA & SSI Youth conference 2012. University of Tokyo, 2012.

Manrique Ruiz Luis Carlos and Yamamoto Kayoko. Compact city analysis in northern and eastern areas of Japan. 20th Annual Conference of GIS Association of Japan. Kagoshima University, 2011.

Other documents non related to this study:

Journal papers

Seki Yoichi, Manrique Ruiz Luis Carlos, Amagai Kenji, Takarada Takayuki. Evaluation of Micro EV's Spreading to Local Community by Multinomial Logit Model. Industrial Engineering & Management Systems. Vol.11, No.2, 2012, pp.148-154.

Jimenez, J.F.P. Soboll, D.S. Manrique Ruiz, Luis Carlos. Bernal, F.E.M. Viera A.M.C. Stereotyped behaviour of ostriches (Struthio camelus domesticus) reared in pairs or trios and monitored during the breeding break. A Hora Veterinaria- Ano 29. No.169, 2009, pp.55-58.

Manrique Ruiz, Luis Carlos. Gutierrez, Edgar. Gonzalez, Gloria. Computable General Equilibrium (EGC) for the Production of Sugar and Bioethanol in Colombia. Engineering Journal Los Andes Univer-

sity. No. 28, 2008, pp.106-112.

CURRICULUM VITAE

Manrique Ruiz Luis Carlos was born in Bogota, Colombia in 1980. He received his bachelor degree in industrial engineering at La Sabana University, Colombia in 2007 where he graduated with honors. In March 2011, he received his Master of Science in information sciences focused on data mining from the Engineering Department of Gunma University, Kiryu, Japan.
Currently, he is enrolled at the University of Electro-Communications in Tokyo, Japan. He is pursuing his doctoral degree in the Department of Social Intelligence and Informatics focused on information systems and his graduation is expected to be in September of 2014. His research interests are spatial data mining, text-mining and real time GIS.

colophon

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

DECLARATION

I Manrique Ruiz Luis Carlos hereby declare that the work entitled Proposal of an evaluation method of a compact city model is my original work. Wherever contributions of others are involved, every effort is made to indicate this clearly with due reference to the literature or acknowledgement, nowhere has any part been written for me by another person.

Tokyo, Japan, September 2014

Manrique Ruiz Luis Carlos